MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE MOULOUD MAMMERI, TIZI-OUZOU

FACULTE DE GENIE ELECTRIQUE ET D'INFORMATIQUE
DEPARTEMENT AUTOMATIQUE

MEMOIRE DE MAGISTER

en ***Automatique***

Option : **Traitement d'Images et Reconnaissance de Formes**

présenté par

ZAABOT Zohra

Ingénieur UMMTO

Les Réseaux Bayésiens. Application en Reconnaissance de Formes à partir d'Informations Complètes ou Incomplètes.

Mémoire soutenu le: / /2012 devant le jury d'examen composé de :

BENFDILA Arezki	Professeur à l'U.M.M.T.O	**Président**
DIAF Moussa	Professeur à l'U.M.M.T.O	**Rapporteur**
HAMMOUCHE Kamal	M.C.A à l'U.M.M.T.O	**Examinateur**
KROUCHI Fazia	M.C.A à l'U.M.M.T.O	**Examinatrice**

Avant propos

Le travail que nous présentons dans ce mémoire a été effectué au sein de laboratoire « Robotique et Vision » du département automatique, faculté de Génie Electrique et d'Informatique, université Mouloud Mammeri, Tizi-Ouzou.

J'exprime mes remerciements et ma gratitude à mon directeur de mémoire, Monsieur DIAF Moussa, professeur à l'Université Mouloud Mammeri, pour l'aide et le soutien et les conseils qu'il m'apportés tout au long de ce travail.

Nos vifs remerciements vont aussi à Monsieur BENFDILA Arezki professeur à l'Université Mouloud Mammeri, pour avoir fait l'honneur de présider le jury de ce mémoire et l'intérêt qu'il a apporté à ce travail.

Nous tenons à remercier Monsieur HAMMOUCHE Kamal, maître de conférences classe A à l'Université Mouloud Mammeri pour avoir bien voulu faire partie de jury de ce mémoire.

Nous remercions aussi Mademoiselle KROUCHI Fazia, maître de conférences classe A à la faculté des Sciences Agronomiques, l'Université Mouloud Mammeri pour avoir accepté de faire partie du jury de ce mémoire, d'avoir mis à ma disposition le pollen utilisé dans ce mémoire et pour tous les conseils qu'elle m'a apportés tout au long de ce travail.

Nous tenons également à remercier Monsieur ABDIOUENE Ahmed chef de station de l' INRF d'El- Kala pour nous avoir fourni le pollen de l'espèce maritime.

Nos remerciements vont aussi à Monsieur AMIROUCHE Saïfi responsable du Microscope électronique à balayage de l'Université Mouloud Mammeri, Tizi Ouzou.

Sommaire

Introduction générale

Chapitre 1: Généralités sur la reconnaissance de formes

1.Introduction.....6
2. Traitement des images.....7
3. Caractérisation de l'objet.....8
3.1. paramètres de formes.....9
3.2. paramètres de couleur.....9
3.3. paramètres de texture.....11
4. Normalisation et réduction des paramètres.....14
5. Classification automatique.....15
6. Applications de la reconnaissance de formes.....17
7. Conclusion.....18

Chapitre 2 : Réseaux bayésiens

1. Introduction.....20
2. Définition d'un réseau bayésien.....21
3. Différents modèles graphiques des réseaux bayésiens.....21
4. Apprentissage de paramètres et de structure.....24
4.1. Apprentissage de paramètres.....24
4.2. Apprentissage de structure.....26
5. Inférences.....28
6. Conclusion.....30

Chapitre 3 : Application des réseaux bayésiens à la reconnaissance automatique du pollen

1. Introduction.....31
2. Généralités sur le pollen 32
3. Acquisition des images.....33
4. Prétraitement des images.....34
5. Extraction des paramètres.....37
5.1. Calcul des paramètres de formes.....37
5.2. Calcul des paramètres de texture.....38
6. Organisation des données.....38
7. Reconnaissance utilisant les paramètres de forme 40
7.1. Utilisation du réseau bayésien naïf 40
7.2. Utilisation du réseau bayésien naïf augmenté par un arbre.....42
7.3. Utilisation du réseau bayésien naïf augmenté par une foret 46
8. Reconnaissance à partir des paramètres de texture.....49
8.1. Utilisation du réseau bayésien naïf 49
8.2. Utilisation du réseau bayésien naïf augmenté par un arbre.....50
8.3. Utilisation du réseau bayésien naïf augmenté par une foret 50
9. Association des paramètres de forme et de texture.....51
9.1. Utilisation du réseau bayésien naïf 51
9.2. Utilisation du réseau bayésien naïf augmenté par un arbre.....52
9.3. Utilisation du réseau bayésien naïf augmenté par une foret 53
10. Récapitulation des résultats et comparaison.....54
11. Conclusion..... 56

Conclusion générale et perspective

Bibliographie

Introduction générale

Dans le domaine de la reconnaissance de formes, on cherche toujours à concevoir et à développer des systèmes intelligents qui peuvent apprendre et raisonner. Cependant, pour des raisons de mauvaises conditions d'acquisition, de transmission et de visualisation des données, les données à traiter ne sont pas toujours complètes et certaines. De plus, il est souvent intéressant d'avoir un ou plusieurs modèles effectuant le lien entre les observations et la réalité pour un objectif précis, et cela, même lorsque les observations sont incomplètes ou imprécises. De ce fait, plusieurs approches probabilistes, parmi lesquelles les réseaux bayésiens, ont été proposées. Ces réseaux conçus pour pouvoir prendre en charge les problèmes comportant la notion d'incertitude, ils apportent, à la fois une interface intuitive sous la forme d'un graphe orienté et un ensemble de méthodes permettant d'exploiter au mieux la connaissance extraite qu'ils modélisent. Ainsi, les réseaux bayésiens sont imposés parmi les différents modèles probabilistes existants en jumelant la théorie des probabilités et celle des graphes. Entre l'ensemble des modèles proposés pour les réseaux bayésiens, notre attention a été portée sur les modèles orientés car ils sont plus intuitifs et plus visuels et dès lors les orientations guident sur la manière d'interpréter la structure, tout comme elles guident lors de l'inférence pour savoir comment faire circuler l'information. Cependant, pour qu'un problème soit résolu efficacement, il faut que le modèle soit correctement construit. C'est ainsi que certaines méthodes d'apprentissage automatique de ces modèles probabilistes ont été développées à partir de connaissances expertes et d'un historique ou d'une base d'exemples du problème à résoudre. La base de données est discrète donc des variables discrètes dans la majorité des

recherches et des études faites sur les réseaux bayésiens. Dans ce présent mémoire, il s'agit pour nous d'appliquer les réseaux bayésiens au domaine de la reconnaissance de formes et plus particulièrement l'identification automatique de grains de ressemblance très grande, de pollen de certaines familles conifères. Ainsi, en premier lieu, l'étude plus ou moins détaillée sur les réseaux bayésiens est nécessaire pour faire ressortir les algorithmes les plus fréquents et choisir une méthode appropriée pour notre application. C'est ainsi que, parmi les variantes des réseaux bayésiens appelées réseaux naïfs ou encore réseaux naïfs augmentés par un arbre où par forêt sont au centre d'intérêt de notre étude.

Ce mémoire a été repartit en trois chapitres. Dans le premier chapitre, nous donnerons un aperçu sur la reconnaissance de forme puisqu'il s'agit du domaine de notre application.

Le second chapitre sera consacré aux réseaux bayésiens et leurs différents algorithmes d'inférence et d'apprentissage à savoir les réseaux bayésien naïfs et les réseaux bayésiens augmentés par un arbre et par forêt (FAN). Cette étude nous permettra de choisir les réseaux bayésiens qui conviennent à notre application, à la seule remarque nos données ne sont pas manquantes mais toujours complètes.

Dans le troisième chapitre, nous présenterons les données images du pollen étudié dans notre application. A partir de ces images, seront calculées les différentes caractéristiques de chaque espèce. L'application de trois modèles de réseaux bayésiens et l'utilisation successives de différentes types de caractéristiques nous permettra d'effectuer une évaluation des résultats obtenus et de procéder à une étude comparative des différente structures des réseaux bayésiens utilisés.

Ce travail sera terminé par une conclusion générale et des références bibliographiques.

Premier chapitre

Généralités sur la reconnaissance de formes

1. Introduction

La reconnaissance des formes peut être définie comme étant l'ensemble des techniques qui permettent de reproduire la perception humaine. Elle consiste à affecter une forme à une classe avec laquelle cette forme présente des caractéristiques semblables. La forme saisie à l'aide d'un capteur approprié est transmise au calculateur où elle est traitée puis classée moyennant des techniques appropriées. Le principe général de la reconnaissance des formes est basé sur différentes opérations à savoir, la localisation de l'objet et l'acquisition de son image à l'aide du capteur approprié dont le manque de qualité nécessiterait un prétraitement de cette image, la représentation des informations acquises et l'extraction de paramètres pertinents et leur réduction pour rendre les donnée plus classifiables, le partitionnement de l'espace de représentation en différentes classes et l'élaboration d'un dictionnaire de référence et en dernier lieu, le classement ou décision qui permet de situer la forme captée dans la classe appropriée. Ainsi, chaque objet est représenté par ces paramètres arrangés sous forme d'un vecteur ligne à m dimensions ou m représente le nombre de paramètres. En effectuant ces calculs pour un ensemble de n objets, nous obtenons donc un tableau de dimension $n_x m$. Cependant, l'information apportée par certaines de ces paramètres peut s'avérer peu

importante. Il est donc nécessaire de procéder au bon choix de ces paramètres ou paramètres pertinents pour atteindre un grand pouvoir de discrimination en faisant appel aux méthodes automatiques de réduction de paramètres dont la plus connue est l'Analyse en Composantes Principales.

Dans ce chapitre nous allons procéder à de brefs rappels sur la reconnaissance de formes qui fait l'objet de ce mémoire tout en insistant sur les parties qui feront l'objet de notre propre application. Ainsi, dans la deuxième partie, nous montrons les différentes manières de représenter les objets et tout particulièrement les paramètres de forme et de texture. Nous ferons ensuite un rappel sur les méthodes de classification qui entrent souvent dans le processus de reconnaissance de formes ainsi que les approches utilisées.

2. Traitement des images

La reconnaissance de formes s'effectue dans la majorité des cas à partir de l'acquisition des images des objets à reconnaître. Ces images sont prétraitées et lissées. Les opérations de segmentation et d'extraction de contours sont parmi les plus importantes. La réduction du temps d'extraction des contours fait encore aujourd'hui l'objet de beaucoup de recherches tout en tenant compte de la présence des bruits et des ombres. Dans une image, les paramètres représentent l'intensité lumineuse très liée aux conditions d'éclairage, la couleur utilisée surtout dans les images de télédétection, la réflectance qui est le rapport entre l'intensité lumineuse réfléchie par un objet et son intensité lumineuse reçue, etc. Si la segmentation consiste à séparer toutes les régions homogènes, la binarisation permet d'extraire les objets de leur fond d'une manière plus rapide que lorsque l'on applique les opérateurs différentiels tels que le Gradient, le Laplacien etc.

Le principe de la binarisation est le suivant : soit L= {0,1,2,...,l-1} l'ensemble des niveaux de gris, l(x,y) la luminance d'un point de cordonnées (x,y), t le seuil permettant de binariser une image et t^* le seuil optimum par rapport à un critère déterminé. Ainsi, dans sa forme la plus simple, la binarisation

consiste, pour une image comportant des objets clairs sur un fond sombre, à choisir un niveau de gris t et à affecter à chaque pixel le label objet ou fond, conformément au test suivant:

- si l(x,y) > t, les pixels sont pris comme étant des objets.
- si l(x,y) < t, les pixels sont pris comme étant le fond.

A partir de l'image binarisée, des paramètres morphologiques autres que texturaux sont extraits.

Il existe plusieurs méthodes de binarisation basées sur le seuillage global lorsque le seuil est déterminé à partir de l'ensemble des niveaux de gris de l'image toute entière de local si le seuil est calculé en tenant compte des propriétés locales de chaque point. Parmi ces deux types de méthodes, on peut citer celles utilisant la maximisation d'entropie [1], la maximisation de la variance interclasse [2], la minimisation de l'erreur quadratique [3] et la préservation des moments pour les méthodes globales et parmi les méthodes de seuillage locales, on peut citer celle qui est basée sur la détection de seuils locaux avec test de validation [4], celle maximisant un critère d'homogénéité [5], et particulièrement la méthode de Otsu [6].

3. Caractérisation de l'objet

On préfère alors décrire les objets par un certain nombre *m* de caractéristiques qui leurs sont propres. Ainsi, chaque objet X_i , *i=1,2,3...,n* est ainsi désigné par un vecteur à *m* composantes tel que:

$$X_i = \left(x_{i1}, x_{i2}, ..., x_{ij}, ..., x_{im}\right) \qquad (1.1)$$

C'est l'ensemble des composantes de ce vecteur, résultats de mesures effectuées sur chaque objet, qui constitue l'espace de représentation. Dans la majorité des cas, c'est sous forme d'un tableau individu/caractère que les observations et les mesures sont rangées (Tableau 1.1). Les paramètres caractéristiques d'un objet saisi par caméra peuvent être de forme, de couleur et de texture.

Tableau 1.1. Tableau individu/caractère

		caractères
		x_1 x_2 ……x_J……x_m
individus	X_1	x_{11} x_{12}……x_{1j}……x_{1m}
	X_2	x_{21} x_{22}……x_{2j}……x_{2m}
	.	.. …… …… . …… .
	.	.. .. …… .. …… ….
	X_i	x_{i1} x_{i2}…… x_{ij}……x_{im}
	.	.. .. …… .. …… ……
	.	.. .. …… .. …… ……
	X_n	x_{n1} x_{n2}……x_{nj}……x_{nm}

3.1 Paramètres de forme

Il existe plusieurs méthodes qui permettent de caractériser la forme de l'objet. Parmi ces paramètres on peut citer la surface, le périmètre, les axes, la circularité, les intersections de lignes, les trous, la compacité, l'élongation, la rectangularité, l'orientation etc. appelées caractéristiques géométriques, ou encore des mesures basées sur les pixels de l'image, appelées descripteurs locaux, ou bien en calculant les moments géométriques telle que les moments invariants [7], ou encore les moments de Zernike [8] tels les suivants :

$$\Phi_1 = \eta_{20} + \eta_{02} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.2)$$

$$\Phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \ldots\ldots\ldots\ldots\ldots \quad (1.3)$$

$$\Phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \ldots\ldots\ldots . \quad (1.4)$$

$$\Phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \ldots\ldots\ldots \quad (1.5)$$

$$\Phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{3})[3(\eta_{21} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (1.6)$$

$$\Phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (1.7)$$

Dans certains cas, les méthodes basées sur les paramètres de formes ne donnent pas des résultats appréciables pour discriminer des objets de formes presque identiques. Dans de telles situations, l'utilisateur peut avoir recours à d'autres types de paramètres comme ceux de la couleur.

3.2 Paramètres de couleur

Une méthode plus courante et efficace consiste à caractériser une image par la répartition des couleurs des différents pixels. Ainsi, la couleur

est un attribut important en reconnaissance de forme. Ils existent plusieurs types de paramètres de couleurs selon l'espace de représentation [9]. Parmi les plus utilisés on peut citer ceux représentés dans les différents espaces. L'espace de base est le RGB (Red, Green et Blue) qui a été largement utilisé grâce à la grande disponibilité d'images dans ce format. En effet, La gamme infinie des couleurs naturelles peut être reproduite à partir de ces trois couleurs seulement. Cet espace propose donc, pour chaque pixel, 256 intensités de rouge, 256 intensités de vert et 256 intensités de bleu soit le pixel est codé sur huit bits, soit chaque pixel est codé sur 24 bits. Il peut aussi être codé sur 32 ou 64 bits. Ainsi pour un pixel, les 24 bits d'une couleur se décomposent en 3 fois 8bits à savoir, 8 bits consacrés à la teinte primaire rouge, 8 bits à la teinte primaire vert et 8 bits à la teinte primaire bleu, ce qui correspond à 256 niveaux différents. Chaque couleur naturelle peut donc être représentée par un système de coordonnées orthogonales à trois dimensions. Ce modèle présente quelques inconvénients : il ne tient pas compte des particularités de la perception visuelle des couleurs (les trois composantes RGB ne permettent pas de reconstituer réellement toutes les couleurs perceptibles par l'oeil humain). Afin de pallier ces inconvénients, la CIE (Commission Internationale de l'Eclairage) a défini un espace de représentation de la couleur basée sur trois couleurs primaires non visibles (dites virtuelles) X, Y et Z. Le passage de l'espace RGB à l'espace XYZ s'effectue simplement grâce à une transformation linéaire pouvant être interprétée comme un changement de base, telle que toutes les couleurs du spectre visible soient contenues dans le triangle XYZ. Ce système présente lui aussi des inconvénients : certaines dimensions descriptives de couleur (clair/foncé, pur/délavé) ne sont pas accessibles directement. Afin de caractériser les couleurs de façon plus intuitive, conformément à la perception naturelle des couleurs, l'espace HSV a été proposé. Celui-ci fait intervenir des critères psychophysiologiques, définis par un cylindre représentant la teinte H (de l'anglais Hue), caractérise la couleur elle même; la saturation S (de l'anglais Saturation) est le niveau de pureté de la couleur représentée (il vaut 0 pour du noir ou du blanc et est maximum pour une couleur pure) ; la valeur V (de l'anglais Value) est la

contenance relative de noir et de blanc. Cependant, ce modèle présent aussi des inconvénients : la teinte n'est pas significative pour les régions peu saturées, très claires, ou très sombres. De plus, le modèle HSV n'offre pas de corrélation simple avec les autres modèles. D'autres espaces de couleur tels le CMJN utilisé pour l'impression, ou le codage TSL (HSL en anglais) destiné aux opérateurs humains, ont été proposés comme on peut les retrouver dans la référence [10]. De toute façon, quel que soit l'espace de représentation utilisé, l'information couleur d'une image peut être représentée par un seul histogramme 3D ou 3 histogrammes 1D. Ces modes de représentation de la couleur ont l'avantage d'être invariants à la translation et à la rotation [11]. Comme dans le cas précédant, pour certains objets, la couleur peut ne pas être suffisante pour une bonne discrimination. Dans de telles situations, l'utilisation d'autres méthodes basées sur des paramètres de texture peut fournir un taux de reconnaissance suffisant.

3.3 Paramètres de texture

Il n'existe pas de définition universelle de la texture. Certains la définissent comme la répétition spatiale d'un motif de base dans différentes directions de l'espace [12]. Cependant, comme pour les notions de forme et de couleur, la notion de texture est liée à la perception humaine. En effet, la texture désigne une surface, représentée par une image ou région d'image, offrant la possibilité de simuler l'apparence tactile de celle-ci. La notion de texture est donc liée aux sens du toucher et de la vision. Une image dite texturée permet de donner la sensation de toucher l'image, de ressentir son grain, ses aspérités, comme si l'on passait son doigt dessus. A partir de cette constatation, on peut définir deux types de textures. En effet, on peut considérer la texture comme la répétition spatiale d'un motif de base dans différentes directions de l'espace. C'est le cas, par exemple, de la texture d'un tissu, d'un mur de briques, d'écailles de poisson, d'un champ labouré, etc. On peut aussi penser qu'une texture ne possède pas de contours francs, mais plutôt un certain désordre, c'est-à-dire une disposition aléatoire que l'on pourrait considérer comme visuellement homogène. On aurait alors une surface fermée, sans motifs isolables ou

répétitifs. C'est le cas, par exemple, de photographies à distance d'herbe, de sable, de graviers etc. Des méthodes statistiques tentent de modéliser les notions qualitatives usuelles de texture: granularité, contraste, homogénéité, répétitivité, fragmentation, orientation, etc. Certaines utilisent les statistiques locales de l'intensité lumineuse tels que les moments d'ordre 1 (histogrammes 1D, moyenne), d'ordre 2 (matrices de co-occurrences, fonctions d'auto-correlation) ou des statistiques d'ordre supérieur (probabilités conditionnelles). La plupart des techniques utilisent les paramètres statistiques du premier ou du second ordre qui ont l'avantage d'avoir un temps de calcul réduit. Les paramètres qui caractérisent la texture sont nombreux et diffèrent selon la méthode utilisée. Parmi ces méthodes, on peut citer la méthode de la dépendance spatiale des niveaux de gris [13], la méthode de la densité spectrale [14], la méthode des longueurs d'iso-segments de niveau de gris [15] et la méthode des différences de niveau de gris [16]. Dans la Méthode de la Densité Spectrale (PSM: Power Spectral Method), le spectre de puissance d'une image x(k,l) est calculé en utilisant la transformée de Fourrier de cette image. A partir du spectre de puissance de l'image, plusieurs paramètres sont calculés. Dans la Méthode de Différence de Niveau de Gris (GLDM:Gray Level Différence Method), des paramètres de texture sont calculés à partir de la matrice de différence de niveau de gris entre les pixels de l'image. Dans la Méthode des Longueurs d'Iso-segments de Niveau de Gris (GLRLM: Gray Level Run Lenght Method), on détecte les suites co-linéaires de points ayant le même niveau de gris dans une direction donnée. La longueur d'un iso-segment est le nombre de points contenus dans l'iso-segment. A partir de la matrice (nombre d'iso-segments, longueur des iso-segments), on peut calculer des cinq paramètres de texture. Parmi ces méthodes, la plus employée est certainement la Méthode de la Dépendance Spatiale des Niveaux de Gris (SGLDM: Spatial Gray Level Dépendance Method) qui est basée sur la notion de matrice de cooccurrence de niveaux de gris. Si I(x,y) est la fonction image définie sur un domaine D à N_g niveaux de gris et telle que I(x,y)=0,1,2,...,N_g et si R est une relation spatiale liant deux pixels, la matrice de cooccurrence relative à la relation R est définie par l'expression ci-dessous où M(i,j) est le nombre de

couples de points du domaine *D* liés par la relation *R*, l'un ayant le niveau de gris *i*, l'autre le niveau de gris *j* .

$$M(i,j) = \{[(x,y),(x',y')] \in D, \quad \text{vérifiant} \quad R\ /I(x,y) = i, I(x',y') = j\} \quad i = 0,1,...,N_g - 1 \quad j = 0,1,...,N_g - 1 \tag{1.8}$$

Divisé par le nombre total de couples, M(i,j) devient une probabilité telle que:

$$m(i,j) = \frac{M(i,j)}{Card\{[(x,y),(x',y')] \in D \text{ verifiant } R\ \}} \tag{1.9}$$

Notons que les couples de points sont définis par la distance *d* entre ces deux points et l'orientation de la droite qui les joint par rapport à l'horizontale. Dans la pratique, la distance *d* est prise égale à 1 et la direction correspond aux quatre orientations à quarante cinq degrés par rapport à l'horizontale. De la matrice de cooccurrence dont la taille est fonction du nombre N_g de niveaux de gris de l'image, les quatre paramètres texturaux suivants sont calculés:

- le Moment Angulaire du Second ordre (MAS) qui constitue une mesure de l'hétérogénéité de la texture et dont la valeur augmente avec l'homogénéité de la texture:

$$SMA = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} m(i,j)^2 \tag{1.10}$$

- la corrélation qui quantifie la dépendance directionnelle des niveaux de gris et qui atteint ses plus grandes valeurs lorsque la direction est voisine de l'orientation des lignes de texture:

$$COR = \frac{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i - \mu_x)(j - \mu_y) m(i,j)}{\sigma_x \sigma_y} \tag{1.11}$$

où :

$$\mu_x = \sum_{i=0}^{N_g-1} i \sum_{j=0}^{N_g-1} m(i,j)\ , \quad \mu_y = \sum_{i=0}^{N_g-1} j \sum_{j=0}^{N_g-1} m(i,j), \quad \sigma_x = \sum_{i=0}^{N_g-1} (i - \mu_x^2) = \sum_{j=0}^{N_g-1} m(i,j)$$

et $$\sigma_x = \sum_{i=0}^{N_g-1} (i - \mu_y^2) = \sum_{j=0}^{N_g-1} m(i,j)$$

- l'entropie qui permet de caractériser le degré d'organisation de l'image et est minimale pour une texture homogène.

$$ENT = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} m(i,j) \log m(i,j) \quad (1.12)$$

- le moment inverse dont la valeur devient de plus en plus faible lorsque l'on s'éloigne de la diagonale principale. Il est proche de l'unité s'il y a une concentration de points autour de la diagonale principale. Ce paramètre permet de mettre en évidence les plages uniformes de texture:

$$MI = \frac{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} m(i,j)}{1+(i+j)^2} \quad (1.13)$$

Quels que soient les paramètres utilisés pour décrire un objet, la phase de réduction et de normalisation et de classification ou d'apprentissage sont de manière classique toujours utilisées.

4. Normalisation et réduction des paramètres

Sachant que les grandeurs des paramètres considérés peuvent être de natures différentes et de valeurs très faibles ou trop élevées pour certaines par rapport à d'autres, leur normalisation est nécessaire avant l'opération de classification. Pour ce faire, la formule suivante a été appliquée :

$$x_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}}, i = 1,2...,n, j = 1,2,...,m \quad (1.14)$$

Où x_{ij}, x_{min}, x_{max} sont respectivement la valeur du paramètre *j* de l'observation *i*, x_{min} la valeur minimale du paramètre *j* et x_{max} la valeur maximale du paramètre *j*.

Signalons toutefois que la normalisation peut présenter des effets négatifs sur la déformation de l'image considérée donc sur la présentation des données. Par ailleurs, le nombre de paramètres extraits initialement est souvent trop élevé d'autant plus que certains de ces paramètres sont pauvres en quantité d'information qu'ils peuvent apporter. L'application de l'Analyse en Composantes Principales [17] nous permettra de réduire ces paramètres aux seuls qui sont significatifs.

5. Classification automatique

Rappelons que la classification automatique est un processus qui permet de regrouper des données dans des ensembles ou classes tel que les éléments d'une classe aient les mêmes caractéristiques alors que la séparation entre les classes doit être totale. Pour réaliser une classification automatique il existe plusieurs méthodes que l'on peut distinguer selon que l'on dispose de toutes les informations *a priori* sur les données. Dans ce cas, la classification automatique est dite supervisée. Dans ce cas, les méthodes utilisées tentent d'apprendre à partir d'images étiquetées (pour lesquelles la classe est connue), constituant un échantillon d'apprentissage, une fonction de classification. Cette fonction permettra d'associer une valeur de la classe de chaque image ou objet non étiqueté (non connu). Dans le cas où l'information n'est pas complète ou sans information *a priori*, on parlera de classification non- supervisée ou clustering.

En reconnaissance de formes, plusieurs méthodes de classification ont été proposées et diffèrent selon le concept utilisé, leurs performances et leurs facultés à discerner des objets malgré leurs proximités. Parmi celles qui sont les plus utilisées, on peut citer la méthode des k-means [18], la méthode hiérarchique ascendant ou descendante [19], celle des k plus proches voisins [20], celle utilisant les réseaux de neurones [21], les réseaux bayésiens [22] et bien d'autres encore [23-24]. La méthode des k-means et ses variantes sont, sans doute les plus populaires et des plus utilisées. Le nom de k-means vient du fait que chacun des k clusters générés est représenté par un centroïde qui n'est autre que la moyenne pondérée des points qui le composent. L'idée initiale est d'utiliser un processus itératif afin de minimiser la somme des carrés des distances entre les exemples et les centroïdes des clusters. Initialement, après la désignation des centres des classes, les distances de tous les points sont calculées et caque point est affecté à la classe qui corresponde à la distance la plus petite. Les nouveaux centroïdes sont calculées et l'opération précédente est ainsi répétée jusqu'à la satisfaction d'un critère ou la stabilisation des classes. Quant aux méthodes hiérarchiques, elles sont ascendantes ou descendantes. La classification hiérarchique consiste à

construire un arbre de classes appelé dendrogramme. Cette construction se fait par deux manières ascendante et descendante. Les algorithmes ascendants ou encore agglomératifs considèrent chaque objet de l'ensemble de données comme des classes initiales et, à chaque étape, on fusionne deux classes qui optimisent un critère de similarité. A l'inverse, les méthodes descendantes ou divisives partent d'une classe unique formée de tous les objets et le divisent petit à petit jusqu'à la satisfaction de critère d'arrêt. Appelée aussi *knn* (k nearest neighbor), la méthode des K plus proches voisins (kppv) consiste à observer les *k plus proches voisins* d'une nouvelle observation afin de décider de la classe d'appartenance de cette nouvelle observation. Pour une nouvelle observation à classer, cet algorithme calcule la distance de cette nouvelle observation à chaque observation pressente dans l'ensemble d'apprentissage. Cette distance est souvent la distance euclidienne alors que d'autres distances sont utilisées telle la distance tangente, distance de Manhattan etc. Ce mécanisme peut devenir extrêmement coûteux en calcul et très demandeur en termes de mémoire de stockage. La méthode de classification basée sur les réseaux de neurones utilise souvent le perceptron multicouche subdivisé en couche de neurones. On nomme généralement la première couche "couche d'entrée" et la dernière couche "couche de sortie". Entre ces deux couches se situent alors une ou plusieurs couches de neurones, nommées couches cachées. Ce type de réseau est très performant pour les taches de classification. Mais, il convient tout de même de préciser qu'il n'y a pas de règles fixes pour la création d'un tel réseau. En effet, bien que le nombre de neurones de la couche d'entrée et de sortie soit imposé par le nombre d'entrées du système (pour la couche d'entrée) ainsi que par la codification des différentes classes (pour la couche de sortie), aucune règle mathématique au sens strict ne permet de déterminer, pour un problème donné, le nombre de couches cachées ainsi que le nombre de neurones de chacune de ces couches. Bien que le calcul soit rapide pour obtenir une réponse du réseau ayant appris, la phase d'apprentissage du réseau peut s'avérer extrêmement coûteuse en ressources de calcul aussi qu'en temps d'exécution. En ce qui concerne les réseaux bayésiens auxquels le second

chapitre leur est consacré, ceux des modèles graphiques dans lequel les connaissances sont représentées sous forme de variables. Chaque variable est un nœud du graphe. Le graphe est toujours dirigé et acyclique (DAG). Il est possible de réaliser des classifieurs performants grâce aux réseaux bayésiens. Il existe plusieurs structures permettant d'employer les réseaux bayésiens comme classifieurs : réseau bayésien naïf, réseaux bayésiens naïfs augmentés par un arbre, réseaux bayésiens semi naïfs condensés etc. L'avantage du raisonnement bayésien est de générer les données manquantes. En outre le résultat de la classification est facilement interprétable. Le choix d'un classifieur est souvent difficile et il se fait en considérant plusieurs critères, une étude comparative a été faite par S.Verron [25]. Le tableau 1.2 montre bien pourquoi on'a opté pour les réseaux bayésiens comme méthode pour la reconnaissance de formes des grains de pollen. Notons aussi que Pernkopf [26] a montré que les classifieurs bayésiens surpassent très souvent les classifieurs a k plus proches voisins, et ce pour des capacités mémoire et des temps de calcul petit. Perzick et al [27] ont comparé les performances d'un simple classifieur naïf de Bayes aux réseaux de neurones et ils ont montré que le plus simple classifieur bayésien (le réseau bayésien naïf) permet dans la plupart des cas d'égaler les réseaux de neurones et même quelques fois de les surpasser. Les auteurs mettent alors l'accent sur les résultats obtenus avec le classifieur bayésien car celui-ci est très simple, demande très peu de ressources de calcul et est très rapide à l'apprentissage.

6. Application de la reconnaissance de formes

Les domaines de la reconnaissance de formes sont très nombreux. On peut notamment citer :

- la reconnaissance de la parole,
- la commande vocale,
- l'identification d'individus,
- la lecture automatique et la lecture optique,
- le diagnostic médical et l'analyse des signaux électro-physiologiques,
- le comptage de globules et la reconnaissance de chromosomes,

- la reconnaissance de signatures,
- l'inspection automatique pour la détection de défauts dans les textiles, les objets industriels,
- l'identification d'objets et le tri d'objets,
- la navigation de robots,

Tableau 1.2 tableau récapitulatif des avantages et inconvénients des différentes méthodes de classification

Classifieurs	SVM	Kppv	Arbres de décision	Réseaux de neurone	Analyse discriminante	Réseaux bayésiens
Gère les variables discrètes	-	+	**++**	-	-	**++**
Gère les variables continues	++	++	+	++	++	++
Gère un grand nombre de classes	-	+	+	+	+	+
Nbre d'observations nécessaires faibles	-	++	+	-	+	+
Temps de calcul admissible	+	-	++	++	++	+
Gère des données manquantes	+	+	+	++	+	++
Gère un grand Nbre de variables	+	+	+	+	+	+
Corrélation entre les variables	++	++	++	++	++	++
Prise en compte du temps	+	+	+	+	+	+
Gère certaines non-linéarités	++	++	++	++	++	++
Facile de réadapter l'outil	+	+	+	+	+	++

++: Signifie que ce classifieur permet de prendre en compte ce critère.
+: Signifie que ce classifieur permet de prendre en compte ce critère mais sous certaines conditions.
-: Signifie que ce classifieur ne permet pas de prendre en compte ce critère.

7. Conclusion

L'existence de plusieurs approches et méthodes de la classification automatique ainsi le nombre de chercheur travaillant dans ce domaine et dans la reconnaissance de formes, montre bien l'importance de ces dernières. Mais malgré leurs efforts, ils ne proposent pas de méthode idéale qui réponde à touts les critères que l'industrie d'aujourd'hui demande. Dans ce chapitre nous avons présenté, de manière générale qu'elles que méthodes les plus usitées dans la classification et on 'a

rassemblé les avantages et les inconvénients de chacune dans un tableau récapitulatif, ce dernier nous a amené à choisir les réseaux bayésiens comme méthode de classification, en répondant mieux aux différents critères cités dessus. L'étude des réseaux bayésiens sera l'objet du deuxième chapitre.

Deuxième chapitre

Réseaux bayésiens

1. Introduction

Un modèle graphique est une famille de distribution de probabilités définie en termes de graphe orienté ou non. Il est constitué de nœuds qui représentent des variables aléatoires et des arcs représentant les relations de dépendances entre ces variables. On distingue deux formes de modèles graphiques probabilistes : les modèles orientés et les modèles non orientés, basés respectivement sur les graphes acycliques orientés connus sous le nom de réseaux bayésiens (RB) [28] et les graphes non orientés, exemple les champs de Markov (HMM)[29] pour ces modèles graphiques probabilistes, non seulement ils bénéficient des avantages des modèles probabilistes, mais aussi ils représentent les avantages liés à leur représentation graphique. En effet ils permettent de visualiser la structure et les propriétés de dépendance conditionnelles du modèle probabiliste correspondant. Ainsi, les RB sont une union entre la théorie des probabilités et la théorie des graphes. Ils constituent une technique d'acquisition, de représentation et de manipulation de connaissance et on les utilise, surtout, pour leur capacité d'effectuer des inférences dans un contexte d'incertitude et aussi pour leurs algorithmes d'apprentissages. Ils sont utilisés pour prévoir, contrôler et simuler le comportement d'un système, à diagnostiquer les causes d'un phénomène observé, à analyser des données et à prendre des décisions. D'ailleurs, les domaines d'applications sont variés. Ils ont été développés pour la première

fois en médecine dans le cadre de la prévention, le diagnostic [30-31], et autre, dans l'industrie, la défense militaire et bien d'autres domaines [32]. Dans le présent travail, les RB sont appliqués à la reconnaissance automatique du pollen.

2. Définition d'un réseau bayésien

Les RB sont des modèles qui permettent de représenter des situations de raisonnement probabilistes basé sur le théorème de Bayés exprimé par la formule (2.1), et ce à partir de connaissances incertaines.

$$P(A/B) = \frac{P(B).P(B/A)}{P(A)} \qquad (2.1)$$

Ainsi, les RB associent une partie qualitative que sont les graphes et une partie quantitative représentant les probabilités conditionnelles associées à chaque noeud du graphe relativement au parent [33]. La partie qualitative exprime des indépendances conditionnelles entre variables et des liens de causalités et ce grâce à un graphe orienté acyclique dont les nœuds correspondent à des variables aléatoires. La partie quantitative est constituée de tables de probabilités dans le cas discret ou de distribution gaussiennes dans le cas continu. Un réseau bayésien $B = \{G, P\}$ est donc défini par un graphe dirigé, un espace probabiliste et un ensemble de variables aléatoires. Le graphe est sans circuit $G = (X, E)$ où X est l'ensemble des nœuds (ou sommets) et E, l'ensemble des arcs. L'espace probabiliste est tel que (Ω, P) où Ω est l'univers des probabilités et P l'ensemble de variables aléatoires X= $\{X_1,...,Xn\}$ associées aux nœuds du graphe et tel que :

$$P(X_1,.., X_n) = \prod_{i=1}^{n} P(X_i / Pa(X_i)) \qquad (2.2)$$

Dans cette expression, $Pa(X_i)$ est l'ensemble des parents du nœud X_i dans G.

3. Différents modèles graphiques des réseaux bayésiens

Il existe plusieurs variantes des RB telles que [34] : les RB multi agents, les RB de niveaux deux, les RB orientés objets, les diagrammes d'influence, les RB dynamiques [35] (temporels), les RB multi entités, les filtres bayésiens [36]:

qui sont des RB dynamiques particuliers et les RB adaptés à la classification tels que ; les RB naïf, les RB naïf augmenté, etc .

En classification, particulièrement, les RB sont largement. Dans ce cas, le nœud parent est considéré comme une variable non observée précisant à quelle classe appartient chaque objet alors que les nœuds enfants sont des variables observées correspondant aux différents attributs caractérisant cet objet. Plusieurs modèles sont conçus dans ce but. Parmi ces réseaux, on peut citer le RB naïf qui est le plus simple, le réseau bayésien augmenté par n'importe qu'elle structure ou par une structure arborescente et autres. Les RB ont une structure simple et unique qui comprend deux niveaux. Le premier niveau contient un seul noeud parent et le second plusieurs enfants avec la forte hypothèse naïve d'indépendance conditionnelle des enfants (X) conditionnellement au parent. Ils sont largement utilisés pour résoudre des problèmes de classification [37]. La figure.2.1 rappelle le principe de fonctionnement et de classification par RBN.

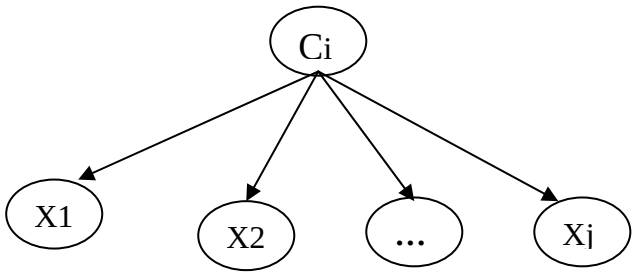


Fig.2.1: Exemple de structure d'un réseau bayésien naïf

Ou :

Ci est le nœud classe et *i* est la ieme classe. *Xj* sont les nœuds des attributs et *j* est le j ieme attribut ou paramètre.

Ce classifieur est connu pour ses performances malgré sa simplicité et il dépasse des techniques beaucoup plus sophistiquées même lorsque l'hypothèse d'indépendance est violée. L'hypothèse d'indépendance des variables permet d'écrire la probabilité a posteriori de chaque classe comme l'indique l'équation suivante :

$$P(C_i / x) = P(C_i)\prod_{j=1}^{p} P(x_j / C_i) \qquad (2.3)$$

Par conséquent, en présence d'un ensemble d'apprentissage, la seule opération à faire est de calculer les probabilités conditionnelles en appliquant la règle de décision « *d* » de Bayes comme suit :

$$d(X) = argmax_{classe}\ P(classe / X)$$

$$= argmax_{classe}\ P(X_j / classe)\ P(classe) \qquad (2.4)$$

$$= argmax_{classe}\ P(C_i)\prod_{j=1}^{p} P\left(X_j / C_i\right)$$

L'hypothèse d'indépendance entre les attributs utilisés dans le RB naïf est généralement fausse (hypothèse naïve). Il existe différentes techniques pour assouplir cette hypothèse [38]. Elles consistent à identifier les dépendances conditionnelles entre les attributs. Nous obtenons alors une sous-structure optimale sur les observations en forme d'arbre en adaptant une des méthodes d'apprentissage de structure, appelée RB augmenté par un arbre (TAN) (fig.2.2).

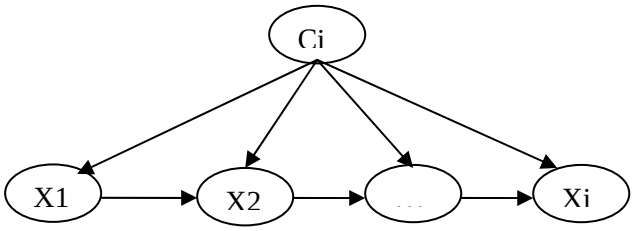


Fig.2.2: Exemple de structure d'un réseau bayésien naïf augmenté

Où *Ci* est le nœud classe,*i* , la ième classe, X_j , les nœuds des attributs et *m,* le nombre d'attributs ou paramètres.

Dont la probabilité à posteriori de chaque classe est l'indiquée par l'équation suivante :

$$P(C_i / x) = P(C_i)\prod_{j=1}^{p} P(x_j^i / Pa(x_j^i), C_i) \qquad (2.5)$$

Où $Pa(x_j^i)$: est le parent de *xj* par rapports à la classe *i* s'il existe.

Notant que des liens inutiles puissent exister dans un TAN. Le nombre des liens est fixé à *n-1* en rajoutant les liens existants entre le nœud classe et les attributs. Parfois, il pourrait être sur-adaptés avec les données. Pour remédier à ce problème certaines conditions sont imposées à la structure TAN pour obtenir une sous structure plus optimale appelée réseau bayésien augmenté par un foret (FAN) [39], dont la probabilité à posteriori de chaque classe et celle

du TAN indiquée par l'équation (2.5). Cependant une application donnée peut être représentée par plusieurs structures et donc différents paramètres ou probabilités, d'où la nécessité d'apprentissage de structure et de paramètre.

4. Apprentissage de paramètres et de structure

La construction d'un réseau bayésien consiste à trouver une structure ou un graphe et estimer les paramètres (probabilités conditionnelles). Cependant devant une très grande base de données, personne ne peut extraire seule la structure adaptée à une telle quantité de données. C'est ici qu'intervient l'apprentissage artificiel.

4.1 Apprentissage des paramètres

L'apprentissage de paramètres consiste à supposer que la structure du réseau et fixe et donc à déterminer les probabilités conditionnelles de chaque variable qui se trouve dans le réseau. Les données disposées peuvent être complètes ou incomplètes, discrètes ou continues. Pour chaque cas, l'algorithme d'apprentissage des paramètres diffère. Dans le cas où toutes les variables sont observées et discrètes, la méthode la plus simple et la plus utilisée pour estimer les paramètres est l'estimation statistique de la probabilité d'un évènement par la fréquence d'apparition de l'évènement dans la base de donnée. Cette méthode est appelée maximum de vraisemblance [40], donnée par l'expression suivante:

$$P(X_i = x_k / \Pi_i = \pi_{ij}) = \hat{\theta}_{ijk}^{MV} = \frac{N_{ijk}}{\sum_{k=1}^{ri} N_{ijk}} \qquad (2.6)$$

N_{ijk} : Le nombre d'occurrences simultanées dans la base de $X_i = x_k$ et $\Pi_i = \pi_{ij}$

avec : $k \in 1.....r_i$ *et* $j \in 1.....q_i$

L'inconvénient de cette méthode est que, on peut avoir une probabilité nulle à cause de la non apparition d'un événement dans la base de données, ce qui est faux en réalité. Pour remédier à ce problème on fait recours à d'autres approches dites méthodes bayésiennes. Ces méthodes sont connues sous le nom de *maximum à posteriori* (MAP) et *Espérance à posteriori* (EAP) [41]. Supposons que les paramètres θ_i admettent une densité de probabilité exponentielle de Dirichlet alors on peut écrire l'expression ci-dessous:

$$P(\theta_i / \alpha_1, \ldots \alpha_n) = \frac{\Gamma\left(\sum_{j=1}^{ri} \alpha_i\right)}{\prod_{i=1}^{ri} \Gamma \alpha_i} \times \prod_{i=1}^{r} \theta^{\alpha_i - 1} \tag{2.7}$$

Avec :

Γ : la fonction gamma d'Euler.

$\Gamma(x+1) = x\Gamma(x)$ et $\Gamma(1) = 1$ dans R

le maximum *a posteriori* est donné par :

$$P(X_i = x_k / \Pi_i = \pi_{ij}) = \hat{\theta}_{ijk}^{MAP} = \frac{N_{ijk} + \alpha_{ijk} - 1}{\sum_{k=1}^{ri} N_{ijk} + \alpha_{ijk} - 1} \tag{2.8}$$

l'espérance *a posteriori* est donné par :

$$P(X_i = x_k / \Pi_i = \pi_{ij}) = \hat{\theta}_{ijk}^{EAP} = \frac{N_{ijk} + \alpha_{ijk}}{\sum_{k=1}^{ri} N_{ijk} + \alpha_{ijk}} \tag{2.9}$$

un autre estimateur appelé estimateur de Laplace [42], qui suppose que les paramètres de Dirichlet égales à 1, est défini par:

$$P(X_i = x_k / \Pi_i = \pi_{ij}) = \hat{\theta}_{ijk}^{EL} = \frac{N_{ijk} + 1}{\sum_{k=1}^{ri} N_{ijk} + 1} \tag{2.10}$$

Dans la plupart des travaux, les calculs des probabilités effectués reposent sur l'utilisation de variables discrètes dont les distributions de probabilité conditionnelles par rapport aux parents sont des tables. L'extension de ces résultats à des variables décrites par des distributions continues est possible. Il est également possible de mixer des variables continues et des variables discrètes comme c'est le cas dans notre application. Pour ce faire, on utilise généralement des distributions gaussiennes et la distribution d'un nœud conditionnellement à ses parents est une distribution gaussienne dont la moyenne est une combinaison linéaire de la valeur des parents et dont la variance et indépendante de la valeur des parents [43].

$$p\left(X_i = \underset{k}{x} \mid \Pi_i = \underset{ij}{\pi}\right) = N(m_i, v_i) \tag{2.11}$$

$$m_i = \mu_i + \sum_{j=1}^{ni} \beta_{ij}\left(x_j - \mu_j\right) \tag{2.12}$$

$$\beta_{ji} = \frac{\sigma_{ij}}{\sigma_j^2} \qquad \text{et} \qquad v_i = \sigma_i^2 \tag{2.13}$$

où :

Πi Sont les parents de *Xi*

μ_i et μ_j sont la moyenne et la variance des l'attributs X_i et X_j respectivement sans prendre en compte les parents.

m_i et v_i : sont la moyenne et la variance de l'attribut X_i en connaissant ses parents X_j .

n_i : est le nombre de parents de X_i.

β_{ij} : appelé matrice de régression des poids ou matrice de poids.

Dans le cas de données incomplets, Il existe plusieurs algorithmes pour estimer les données manquantes à partir des données connues, au lieu de les ignorer. le plus utilisé est l'algorithme EM (Expectation maximisation ou Espérance maximisation) [44] (annexe.1).

4.2. Apprentissage de structure

L'apprentissage de structure ayant pour but de trouver le meilleur réseau permettant de représenter les données le mieux possible. Cependant, la recherche dépend de nombre de variables, d'arcs et de valeurs. Le nombre de structure généré possible à partir de *n* noeuds est très grand, il est donné par la relation suivante :

Soit : $r(n)$ le nombre de graphe possible, et n le nombre de nœuds existants.

$$r(n) = \sum_{i=1}^{n} (-1)^{i+1} C_n^i \, 2^{i(n-i)} \, r(n-i) = n^{2^0(n)} \qquad (2.14)$$

Ainsi pour *n=4* on' a *r (4) =543* et pour *n = 7* on' a *r (7) = 1,4.10⁹.*

Beaucoup de travaux se sont intéressés aux problèmes de l'apprentissage de structures [45]. La aussi comme pour l'apprentissage de paramètres, on a deux cas, selon que les données sont totalement ou partiellement observables. Pour le premier cas deux familles d'approches ont été proposées. A savoir celles qui sont basées sur des scores pour trouver la configuration optimale d'un réseau bayésien. Ces algorithmes consistent à parcourir tous les graphes possibles puis associer un score à chaque graphe. Le graphe qui possède le plus grand score va être sélectionné. Cependant, cette méthode est applicable seulement pour des problèmes à taille limitée (quelques centaines de variables). Il existe

deux types de score : les scores locaux et les scores globaux [46]. La deuxième famille utilise des tests statistiques afin de déterminer les indépendances entre les variables dans le réseau. In peut citer l'algorithme de « PC » pour Peter et Clark, « IC » algorithme de principe similaire, pour Inductive Causation, « BNPC » Nommée BN-PC-B pour Bayes Net Power Constructor et B, car les mêmes auteurs introduisent deux algorithmes, le premier et BN-PC-A, « GS » comme Greedy Search ou recherche gloutonne et la « Méthode de recherche de l'arbre de recouvrement de poids maximal » (Maximal Weight Spanning Tree ou MWST). Cette méthode s'applique à la recherche de structure d'un réseau bayésien en fixant un poids à chaque arête potentielle X_i–X_j de l'arbre [47]. Ce poids peut être un score ou une information mutuelle [48] entre les variables X_i et X_j, qui est donnée en cas de variables discrètes par l'expression suivante:

$$I(X;Y)=\sum_{x\in X}\sum_{y\in Y}P(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{2.15}$$

Et elle donnée en cas de variables continues par:

$$I(X_i;X_j)=-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\log\left(1-\rho^2\left(X_i,X_j\right)\right) \tag{2.16}$$

Avec :

$$\rho\left(X_i,X_j\right)=\frac{\sigma_{ij}}{\sqrt{\sigma_i^2\sigma_j^2}} \tag{2.17}$$

Ou :

$\rho\left(X_i,X_j\right)$: Appelé cœfficients de corrélation entre X_i et X_j ou matrice de poids.

Cette matrice de poids une fois définie, il suffit d'utiliser un des algorithmes de résolution du problème de l'arbre de poids maximal comme l'algorithme de Prim ou celui de Kruskal [49]. Le principe de l'algorithme de Kruskal pour trouver un arbre de poids minimum dans un graphe *G* est de trier les arcs par ordre croissant de leur poids, puis dans cet ordre, les arcs sont ajoutés un par un dans un graphe *G'* pour construire progressivement l'arbre. Un arc est ajouté seulement si son ajout dans *G'* n'introduit pas de cycle, autrement dit, si *G'* reste un arbre. Sinon, on passe à l'arc suivant dans l'ordre du tri. Contrairement à l'algorithme de Prim le tri des arcs se fait par ordre décroissant. L'algorithme de Kruskal est donné par la figure (2.3):

Algorithme de Kruskal

Entrées: G = (X; U) un graphe.
Sortie: U' un ensemble d'arcs.
Variables intermédiaires: i un entier.

Début
Trier les arcs de G dans l'ordre croissant des poids;
[On les notera u1, u2,..., um.]
$U' \leftarrow \phi i$;
Pour i $\leftarrow$ 1 à m faire
Si le graphe (X; U' $\cup$ { ui }) ne contient pas de cycle alors
U' $\leftarrow$ U' $\cup$ { ui };
Fin si;
Fin tant que;
Fin

Fig 2.3 le principe de l'algorithme de Kruskal

Puis l'arbre non dirigé retourné par cet algorithme doit ensuite être dirigé en choisissant une racine et en parcourant l'arbre par une recherche en profondeur [50]. La racine peut être choisie aléatoirement ou à l'aide de connaissance *a priori*. Comme dans l'apprentissage de paramètres, si nous ne possédons pas toutes les mesures qui sont issues des observations réelles, on utilise les algorithmes d'apprentissage de structure avec des données incomplètes, l'algorithme le plus usité est celui de EM (Expectation maximisation ou Espérance maximisation) [51].

En présence d'un réseau bayésien, nous pouvons extraire un certain nombre d'informations. En premier lieu, nous avons accès à la structure du réseau, celle-ci nous permet de savoir quels sont les attributs qui sont dépendants ou non, de plus nous avons accès aux probabilités conditionnelles. Le calcul de ces informations constitue l'inférence.

5. Inférences

L'inférence est le calcul de la probabilité de n'importe qu'elle variable d'un modèle probabiliste à partir de l'observation d'une ou de plusieurs autres variables. Il consiste à propager une ou plusieurs informations au sein de ce réseau, pour en déduire comment sont modifiées les croyances concernant les autres nœuds. La structure du graphe joue un rôle important dans la complexité de ces calculs ainsi dans le choix de la méthode d'inférence. On peut distinguer deux catégories d'algorithmes d'inférence [52] : l'inférence exacte et l'inférence approché. Plusieurs méthodes ou algorithmes conçues

spécialement pour les problèmes d'inférence exacte pour les réseaux bayésiens. On peut citer l'algorithme de passage de messages de Pearl (annexe 2), dont le principe est le suivant, à chaque nœud est associé un processeur qui peut envoyer des messages à ses voisins, jusqu'a ce qu'un équilibre soit atteint, en un nombre fini d'étapes. Le problème de l'algorithme de passage de messages de Pearl est qu'il ne s'applique qu'aux arbres, l'algorithme d'arbre de jonction, qui est une généralisation de l'algorithme de passage de messages de Pearl permet de faire de l'inférence sur n'importe quel type de graphe. Cette méthode est divisée en cinq étapes qui sont : la moralisation du graphe, la triangulation du graphe moral, la construction de l'arbre de jonction, l'inférence dans l'arbre de jonction en utilisant l'algorithme des messages locaux et la transformation des potentiels de clique en lois conditionnelles mises à jour. A rappeler qu'une clique est un sous graphe du graphe G dont tous les noeuds sont connectés deux à deux. On peut citer encore l'Algorithme d'élimination des variables ou l'algorithme d'élimination de Bucket (annexe 3).qui consiste à marginaliser la distribution de probabilité jointe d'un reseau, en procédant variable par variable. Chaque marginalisation sur une variable *Xi* donne lieu à une somme des probabilités de cette variable. Parfois, cette somme vaudra 1, ce qui conduit à l'élimination de la variable *Xi*. On procédera alors à la marginalisation sur une des variables restantes et ainsi de suite jusqu'a ce que la distribution soit marginalisée. Le problème de cet algorithme est que l'ordre dans lequel les variables sont éliminées détermine la quantité de calcul nécessaire pour marginaliser la distribution de probabilités jointe et donc la complexité de l'algorithme. Dans le cas de la classification on peut avoir soit des variables discrètes ou bien un mélange de variables discrètes représentant les classes et les variables continues représentants les caractéristiques du modèle. Dans les deux cas le problème de l'inférence revient à calculer les probabilités *à posteriori* suivantes, ainsi la classe choisie est celle qui maximise ces probabilités :

$$P(c_i / x) = P(c_i) \prod_{j=1}^{n} P(x_j / Pa(x_j), c_i) \quad \text{Si } X_j \text{ admet un parent.}$$

$$P(c_i / x) = P(c_i) \prod_{j=1}^{n} P(x_j / c_i) \quad \text{Si non.}$$

Avec :

$P(x_j / Pa(x_j), c_i)$: est la fréquence d'apparition de x_j en connaissant $Pa(x_j)$ dans la classe c_i si cas de variables discrètes donnée par l'équation (2.6). Ou bien :

$P(x_j / Pa(x_j), c_i)$: est une distribution gaussienne donnée par l'équation (2.11).

Dans le cas ou les probabilités conditionnelles ne sont pas exactes, pour effectuer une inférence exacte avec ces valeurs approximatives n'est plus probant. Il est alors intéressant d'effectuer une inférence approchée, en utilisant d'autres algorithmes tels que l'Algorithme de Métropolies Hastings [53], l'échantillonneur de Gibbs, Loopy belief propagation [54]…etc.

6. Conclusion

Nous avons illustré dans ce chapitre, qu'est ce qu'un réseau bayésien, et la diversité de leurs domaines d'application. Nous avons constaté que les réseaux Bayésiens possèdent tous les avantages des modèles probabilistes standards, ainsi que des avantages supplémentaires liés à leur représentation graphique. De même, leur représentation facilite la compréhension dans un domaine de connaissances (elle représente directement les connaissances du domaine et non des procédures de raisonnement) et modélise explicitement tous les liens de dépendances entre variables. Nous avons montré que la difficulté dans le réseau bayésien consiste dans la construction de la structure en utilisant les différents algorithmes d'apprentissage de structure, ainsi l'estimation de ses paramètres, que ce soit les tables de probabilités s'il s'agit de données discrètes ou encore des probabilités des distributions gaussiennes s'il s'agit de données ou variables continues. Une fois le réseau est défini, et grâce aux différentes méthodes et algorithmes d'inférences développées spécialement pour les réseaux bayésiens, l'information contenue dans ce réseau peut être calculée en n'importe quel nœud du réseau. De plus nous avons développé quels que modèles adaptés à la classification, tels que les réseaux bayésiens naïf (RBN), le réseau bayésien augmenté par un arbre (TAN) ou encore celui augmenté par foret(FAN). Une application de ces réseaux bayésiens dans la reconnaissance du pollen, fera l'objet du troisième chapitre.

Troisième chapitre

Application des réseaux bayésiens à la reconnaissance automatique du pollen

1. Introduction

Dans le but de mettre en évidence les différents algorithmes d'apprentissage et d'inférence dans les réseaux bayésiens adaptés à la classification et qui ont été discutés précédemment dans le deuxième chapitre, nous avons choisi comme application la détection et la reconnaissance automatique du pollen. Dans cet objectif, plusieurs chercheurs ont mis en évidence différentes méthodes automatiques compte tenu de l'importance des domaines d'application telle l'agriculture, la médecine, l'écologie etc. Ces méthodes sont basées sur l'introduction ou le développement de techniques de traitement d'image microscopiques et de reconnaissance de formes [55] [56]. Dans le cadre de ce mémoire, nous nous sommes intéressés au pollen de pin et de cèdre. Pour cela nous avons prélevé trois espèces de pollen de la même famille, à savoir le pollen du pin d'Alep cueilli le mois de mai à Tikjda, le pollen du pin maritime cueilli le mois de mars passé à El-Kala et le pollen de cèdre cueilli le mois de février passé à Tala-Guilef, versant nord-ouest du Djurdjura. La technique de reconnaissance automatique utilisée est basée sur les réseaux bayésiens et sachant que les images acquises sous microscope électronique à balayage sont en niveaux de gris, seuls les paramètres de forme et de texture sont pris en considération. Ainsi, dans ce chapitre, après une brève présentation sur le pollen, suivront ensuite les étapes d'acquisition et du prétraitement des

images, le calcul des paramètres de forme et de texture, l'application de trois modèles de réseaux bayésiens et l'interprétation et l'évaluation des résultats.

2. Généralités sur le pollen

Le mot pollen dérive du latin et signifie « fine farine », présenté sous forme de poudre jaune dont les grains ne peuvent être visibles qu'à l'aide d'un microscope tant sa taille est de l'ordre de 2.5 µm à 200 µm (fig.3.1).

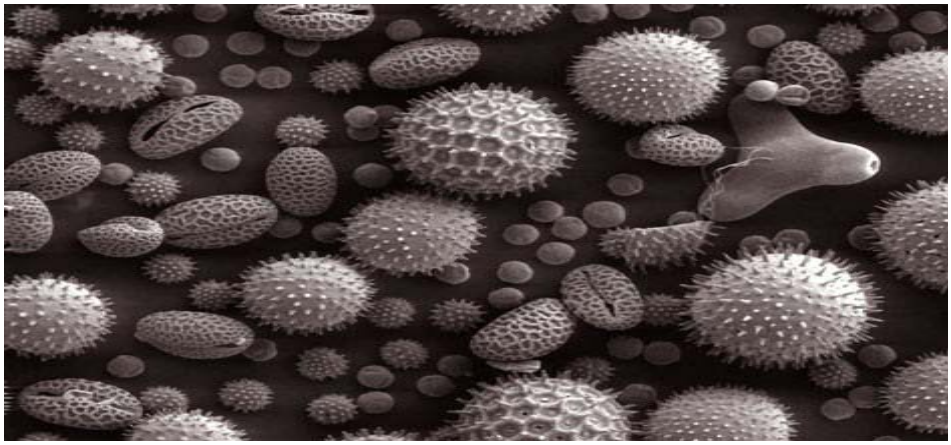


Fig.3.1 : Exemples de pollens vus au microscope électronique.

Le pollen de pin se forme entre les écailles des cônes (fig.3.2).

Cône jeune

Cône mur

Fig.3.2 Cône de pin

Le pollen de pin est libéré à de très fortes températures et est dispersé grâce au déplacement turbulent de masses d'air chaud qui peuvent le transporter sur des distances importantes. Un grain de pollen, contrairement aux cellules, possède deux noyaux: un noyau végétatif et un noyau reproducteur. La figure (3.3) montre le schéma d'un grain de pollen de pin constitué de deux cellules de tailles inégales avec cette particularité que la plus petite est entièrement enfermée dans la plus grande.

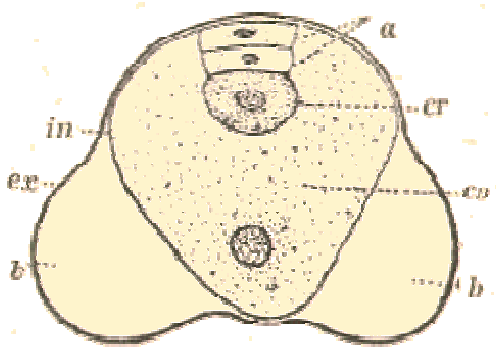


Fig.3.3 : Schéma d'un grain de pollen de Pin sylvestre (Pinus sylvestris) [57].

Cv, cellule végétative. - cr. Cellule génératrice des deux gamètes. - a, deux cellules accessoires. - in et ex, intine et exine. - b et b, deux ballons latéraux remplis d'air.

La plus grande des deux cellules *cv*, qui possède en même temps un noyau plus volumineux, est appelée la cellule mère du prothalle mâle ou encore la cellule végétative parce qu'elle engendrera dans la suite un petit prothalle rudimentaire et transitoire dans lequel se développent deux gamètes mâles. La plus petite des deux cellules, *cr*, qui possède en même temps le plus petit noyau, engendrera dans la suite deux gamètes mâles, ce qui la fait appeler la cellule mère des gamètes ou encore la cellule génératrice.

La cellule génératrice est accompagnée à la périphérie de deux autres petites cellules (*a*). Ce sont des cellules accessoires ou cellules stériles. Le grain de pollen possède à sa périphérie deux membranes : une interne, l'*intine*, qui est mince et une externe, qui s'appelle l'*exine*; chez beaucoup d'espèces (Pin, Mélèze), cette dernière se renfle sur les côtés et forme deux boursouflures remplies d'air (*b, b*) qui allègent le pollen et facilitent sa dissémination par le vent.

3. Acquisition des images

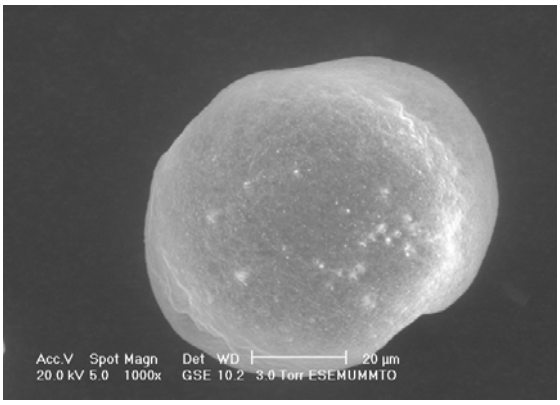
Pour l'acquisition des images de ces pollens, nous avons utilisé le microscope électronique à balayage (MEB) du laboratoire de microscopie de l'université Mouloud Mammeri de Tizi Ouzou (fig.3.4)

Fig.3.4 : photo du microscope électronique à balayage utilisé

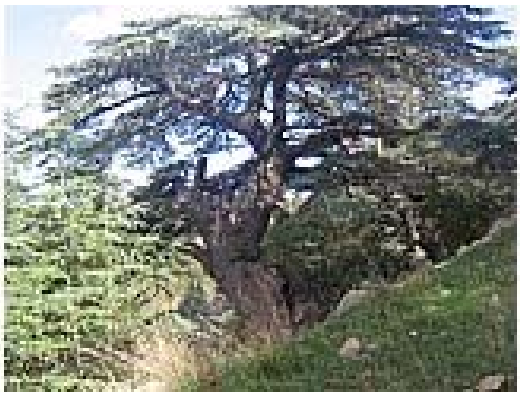
Cent images en niveau de gris de chacune des trois espèces ont été acquises sans aucune préparation préalable avec un agrandissement de 1000 X. La figure (3.5) montre un exemple de ces images.

4. Prétraitement des images

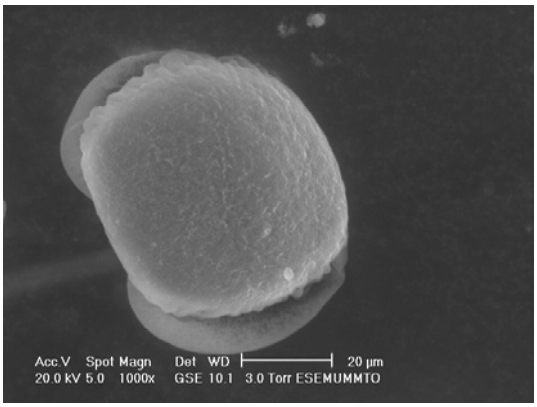
Le prétraitement des images est une étape importante car elle permet de sélectionner, dans l'espace de représentation, l'information nécessaire à l'application. Cette sélection passe par l'élimination du bruit dû aux conditions d'acquisition. Ainsi, par la suppression de la redondance, le nouvel espace de représentation a une dimension réduite. Dans l'image, seule l'information servant à caractériser l'objet est conservée. S'il arrive que l'on ait plus d'une cellule dans l'image, une pré segmentation en utilisant des opérateurs de la morphologie mathématique [58] tels que la dilatation, l'érosion, l'ouverture, la fermeture, le remplissage, est alors appliquée dans Le but de séparer la cellule entière des autres cellules qui peuvent apparaître sur les bords de l'image, c'est-à-dire des cellules incomplètes (fig.3.6).



Pollen de cèdre

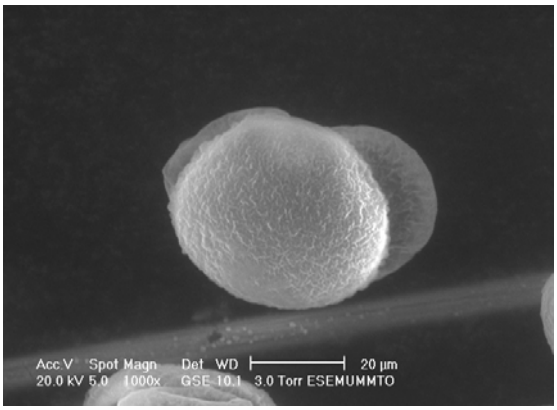


Arbre du cèdre



Pollen de pin maritime

Arbre du pin maritime



Pollen de pin d'Alep

Arbre du pin d'Alep

Fig.3.5 Les trois espèces de pollens considérés et leurs arbres correspondants

Ainsi, les étapes nécessaires que nous avons suivies pour ce prétraitement sont les suivantes [59] :

- Lecture de l'image;
- Détection des contours internes et externes d'une cellule,
- Linéarisation des contours des objets, en utilisant un élément structurant linéaire;
- Dilatation des contours des objets;
- Remplissage interne des objets;
- Elimination des objets dépassant la bordure de l'image;
- Adoucissement de l'objet et segmentation (contour externe seulement) en utilisant l'opération d'érosion.

Image originale

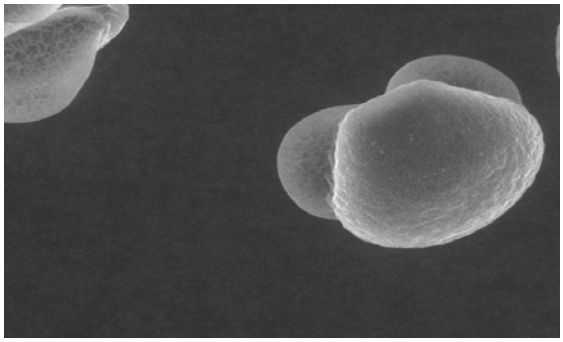
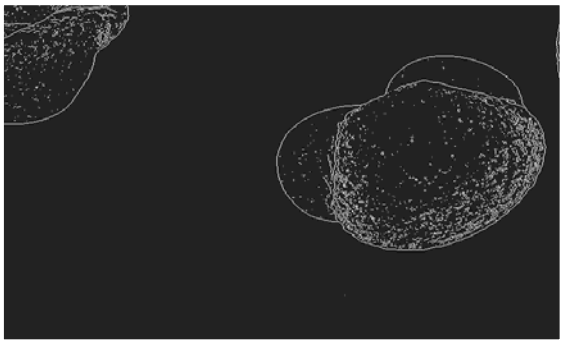


image après traitement



Le contour de l'image

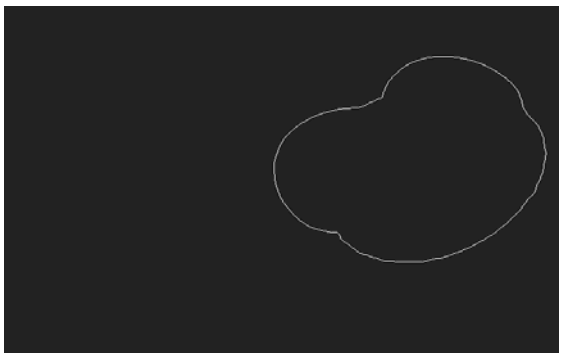


Fig.3.6: différentes étapes de segmentation

Une fois l'image segmentée, un autre traitement consistant à extraire les caractéristiques indispensables à la reconnaissance de formes est effectué. Les caractéristiques doivent d'être invariantes en rotation, en translation et en homothétie.

5. Extraction des paramètres

Etant donné que les images issues du microscope considéré sont en niveaux de gris, nous avons ainsi commencé par considérer d'abord les paramètres de forme puis les paramètres de texture en tenant compte des résultats présentés dans [60] concerne leurs pertinences.

5.1 Calcul des paramètres de forme

Ainsi, le travail [61] réalisé au sein de notre laboratoire « Robotique et Vision », en ce qui concerne la sélection des paramètres où l'analyse en composantes principales a été appliquée, seules quatre caractéristiques géométriques de forme, à savoir la surface S le périmètre P, l'axe majeur AB et l'axe mineur ab ont été sélectionnés. Rappelons que la mesure de ces paramètres est donnée en nombre de pixels. Notons que l'axe majeur AB et l'axe mineur ab sont respectivement calculés à partir du centre de gravité G (X_g, Y_g) et du centre d'inertie I (X_I, Y_I) donnés par les formules suivantes :

$$X_G = \frac{\sum_{i=1}^{N} x_i}{N} \quad \text{et} \quad Y_G = \frac{\sum_{i=1}^{N} y_i}{N} \tag{3.1}$$

$$X_I = \frac{\sum_{S=1}^{NS} x_S}{N} \quad \text{et} \quad Y_I = \frac{\sum_{S=1}^{NS} y_S}{N} \tag{3.2}$$

Le tableau.3.1 donne un exemple de valeurs de ces paramètres pour les trois espèces de pollens étudiés. Dans ce tableau, la première ligne correspond aux paramètres du pin d'Alep, la seconde à ceux du pin maritime et la troisième à ceux du cèdre.

Tableau.3.1 Exemples de valeurs des paramètres de forme.

	S	**P**	**AB**	**ab**
Alp	987.00	1155.58	529.89	404.24
Mar	1053.00	1228.21	546.43	445.83
Ced	1294.00	1522.65	666.21	599.20

5.2 Calcul des paramètres de texture

Etant donné que certains pollens sont proches des autres de par leurs formes, nous ne nous sommes intéressés aussi aux paramètres de texture. Nous considérons les moments d'espace de second ordre issus de la matrice des cooccurrences. Pour ces paramètres sont descriptifs de la texture, il s'agit notamment du contraste, de l'homogénéité, de l'entropie et de la corrélation. Notons que même R.W.Conners et C.A.Harlow dans [62] ont montré que ces quatre paramètres permettent une meilleure discrimination des textures. Le tableau.3.2 donne un exemple de valeurs de ces paramètres calculés pour les trois espèces considérées de pollen.

Tableau.3.2 Exemples de valeurs des paramètres de texture à 45

	Homogénéité	**Contraste**	**Entropie**	**Corrélation**
Alp	0.8833	0.7067	0.5581	0.4736
Mar	0.8543	0.9607	0.4221	0.5451
Ced	0.8583	0.8347	0.3208	0.6263

6. Organisation des données

Afin de classifier un ensemble de grains de différents pollens, l'ensemble de mesures de formes et de texture sont effectués sur toute la base d'images disponibles. La base d'images est alors divisée en deux sous-bases: celle des images pour lesquelles la classe est connue et qui constitue l'échantillon d'apprentissage et celle du reste des images qui sont celles à reconnaître et qui constituent l'échantillon de test. Dans le cadre de notre travail, la base d'apprentissage comporte 150 images réparties uniformément sur trois classes et pour la base de test, nous avons disposé aussi de 150 autres

images réparties sur les trois classes. Dans ces bases, chaque objet est représenté par des vecteurs de caractéristiques qui sont les paramètres de formes puis les paramètres de texture en deuxième lieu et en fin l'association des deux. Les classes sont représentées par un nœud discret qui a comme valeur 1, 2 et 3 représentant respectivement le pin d'Alep, le pin maritime et le cèdre. Dans notre application, nous avons utilisé trois modèles de réseaux bayésiens à savoir les réseaux bayésiens naïfs RBN, le TAN (Réseaux bayésiens augmentés par un arbre) et le FAN (Réseaux bayésiens augmentés par une forêt). Pour ces trois modèles, l'apprentissage des paramètres et de structures est indispensable avant de faire toute inférence sauf pour le réseau bayésien naïf dont la structure est déjà connue. La figure (3.7) illustre l'architecture du système que nous avons utilisé.

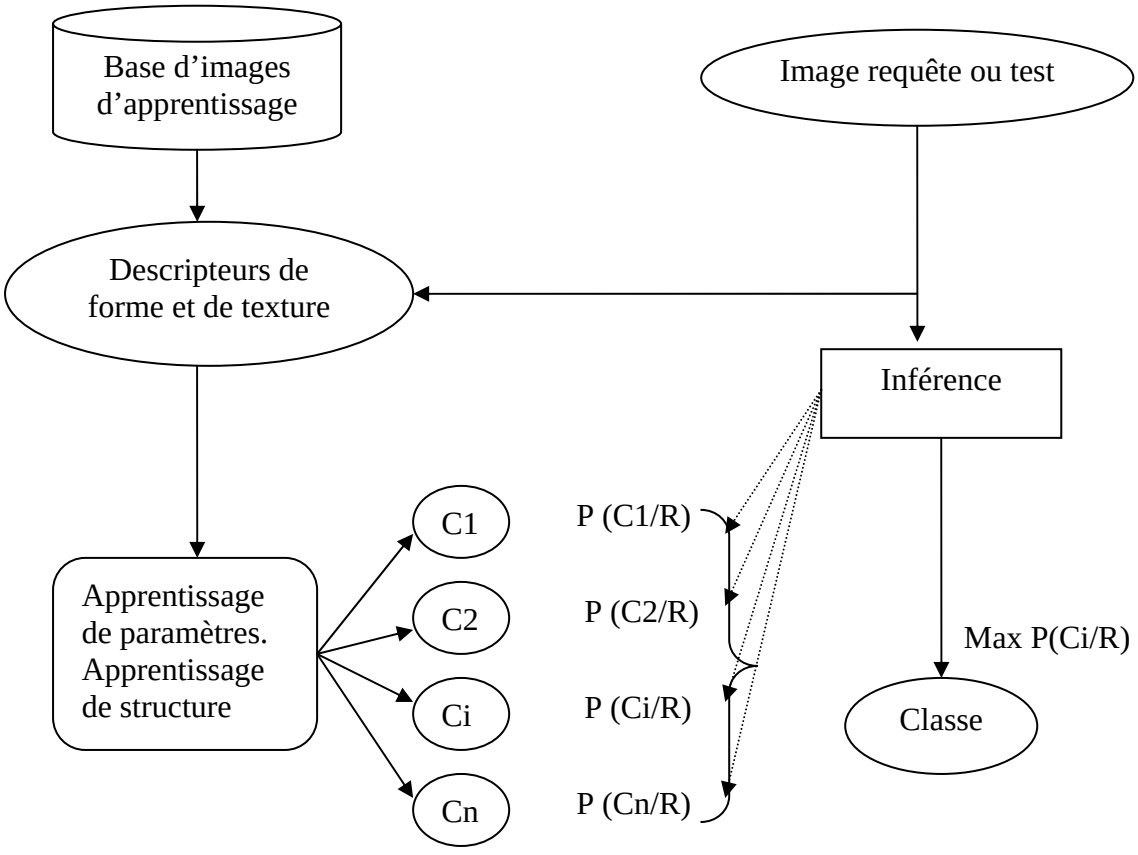


Fig.3.7 Structure du système de reconnaissance utilisé.

7. Reconnaissance utilisant les paramètres de forme

Pour reconnaître les espèces de pollen à partir des paramètres de forme cités précédemment, nous allons utiliser les trois modèles RBN, les TAN et les FAN.

7.1 Utilisation du réseau bayésien naïf

Partant des 150 images des trois espèces de pollen représentant l'ensemble d'apprentissage, le réseau bayésien naïf correspondant à cet ensemble est représenté par la figure (3.8).

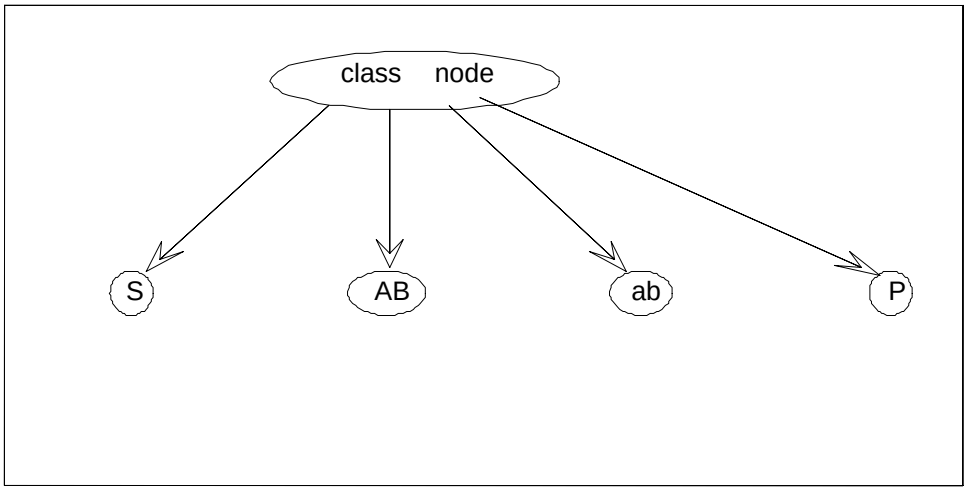


Fig.3.8 : Réseau bayésien naïf obtenu avec paramètres de formes

Notons que les quatre nœuds enfants correspondent aux quatre attributs caractérisant les quatre caractéristiques de forme du pollen, à savoir la surface, le périmètre, l'axe majeur et l'axe mineur. Ces paramètres sont considérés comme indépendants dans le contexte du nœud parent qui correspond à une classe. Notons que cette structure est la plus simple des structures des réseaux bayésiens. L'apprentissage des paramètres du réseau bayésien peut être réalisé selon le principe suivant. Partant du fait que l'ensemble d'apprentissage ou entrée du réseau et la structure du RBN sont connues, il s'agit de déterminer la classe de l'image, ce qui correspond à l'identification du pollen. Pour ce faire, les étapes permettant l'apprentissage des paramètres du RBN, nous avons suivi les étapes suivantes:

- Détermination des fréquences d'apparition de chaque classe à partir de l'ensemble d'apprentissage. Ces fréquences sont calculées par les expressions suivantes :

$$P(c_i) = \frac{N(C_i)}{N} \quad et \quad \sum_{i=1} P(c_i) = 1 \tag{3.3}$$

où N (C_i) représente le nombre d'image ou objet appartenant à la classe C_i et N, le nombre total d'images d'apprentissage.

- Calcul de la probabilité conditionnelle de chaque attribut par rapport à son parent en appliquant l'expression suivante :

$$P(X_j / C_i) = N\left(\mu_j, \sigma j / C_i\right). \tag{3.4}$$

Où N est la loi normale gaussienne.

Ainsi, l'apprentissage de paramètres, dans ce cas revient à calculer les paramètres de la gaussienne représentant la distribution gaussienne de la probabilité conditionnelle de chaque attribut dans la classe C_i . Dans cette expression, μ_j et σ_j sont la moyenne et la variance de l'attribut j dans la classe C_i. Après avoir calculé toutes les moyennes et les variances de tous les attributs de chaque classe, le réseau est ainsi quantifié et il peut être utilisé pour classer de nouveaux objets étant données leurs nouvelles valeurs d'attributs en utilisant l'équation suivante :

$$P(c_i / x) = P(c_i) \prod_{j=1}^{n} P(x_j / c_i) \tag{3.5}$$

Cette équation permet de calculer la probabilité de toutes autres valeurs d'attributs par rapport à chaque classe. L'objet sera affecté à la classe qui correspond la probabilité *P(ci/ x)* est la plus grande. La matrice de confusion en utilisant les paramètres de formes, pour le réseau bayésien naïf, est donnée par le tableau 3.3.

Tableau.3.3 Matrice de confusion en utilisant les paramètres de forme

	Alp	**Mar**	**Ced**	**Rejet**
Alp	**46**	4	0	0
Mar	2	**48**	0	0
Ced	0	5	**45**	0

La matrice de confusion, en utilisant la forme, montre que quatre grains de pollen du pin d'Alep sont confondus avec les grains du pin maritime. Deux gains de pin maritime sont confondus avec le pin d'Alep et cinq grains de pollen du cèdre sont confondus avec le pin maritime.

7.2 Utilisation du réseau bayésien naïf augmenté par un arbre

La structure pour ce modèle de réseau bayésien naïf augmenté par un arbre (TAN) n'étant pas connue, la recherche de la structure adaptée à notre problème est nécessaire. Pour ce faire, nous avons suivi l'algorithme de la figure (3.9).

Algorithme d'apprentissage de la structure :

Entrée :- L'ensemble d'apprentissage.
- RB Naïf

Sortie :- RB Naïf augmenté par un arbre

Algorithme :

1- Calculer l'information mutuelle conditionnelle pour chaque couple d'attributs donnée par :

$$I_P(X_i; X_j \mid C) = -\frac{1}{2}\sum_{c}^{C} P(c)\log\left(1-\rho_c^2\left(X_i, X_j\right)\right)$$

Avec $\rho_c\left(X_i, X_j\right) = \dfrac{\sigma_{ij|c}}{\sqrt{\sigma_{i|c}^2\sigma_{j|c}^2}}$ et $i \neq j$

2- Construire un graphe complet et non orienté entre les attributs ;

3- Pondérer $I_P(X_i; X_j \mid C)$ sur l'arc (X_i, X_j) ;

4- Transformer le graphe en un arbre dont la somme des arcs est maximum en utilisant l'algorithme MWST.

5- Choisir un nœud comme racine de l'arbre et le transformer en un arbre orienté. Nous choisissons l'attribut X_{racine} qui a l'information mutuelle maximale avec la classe, définie par l'équation ci-dessous, comme racine

X_{racine} = arg max IM (X_i, C)

Il est naturel d'utiliser cette stratégie, c'est à dire l'attribut qui à la plus grande influence sur la classification devrait être la racine de l'arbre.

6- Ajouter le nœud C et un arc de C vers chaque nœud X_j.

Fig.3.9 Algorithme d'apprentissage de la structure du TAN

Comme pour le réseau naïf, l'apprentissage des paramètres dans le cas du TAN est basé sur le calcul de fréquences de chaque classe ainsi que des paramètres de la loi normale dans la base d'apprentissage conformément à l'algorithme de la figure (3.10). Pour nos expériences, nous avons utilisé la Bayes Net Toolbox de Murphy [63] et la Structure Learning Package pour

faire l'apprentissage de structure [64]. En effet, en appliquant l'algorithme ci-dessus, la matrice des scores calculés à base de l'information mutuelle donnée par l'expression suivante

$$I_P(X_i; X_j \mid C) = -\frac{1}{2}\sum_{c}^{C} P(c)\log\left(1-\rho_c^2\left(X_i, X_j\right)\right) \tag{3.6}$$

Avec $\rho_c\left(X_i, X_j\right) = \dfrac{\sigma_{ij|c}}{\sqrt{\sigma_{i|c}^2\sigma_{j|c}^2}}$ et $i \neq j$

Est la suivante:

$$IMC = \begin{bmatrix} 0 & 1.1221 & 0.2520 & 2.7264 \\ 1.1221 & 0 & 0.1187 & 1.1080 \\ 0.2520 & 0.1187 & 0 & 0.2686 \\ 2.7264 & 1.1080 & 0.2686 & 0 \end{bmatrix}$$

Algorithme d'apprentissage de paramètres :

Entrée :
- L'ensemble d'apprentissage.
- Structure des réseaux bayésiens naïf augmenté par un arbre.

Sortie : les paramètres des réseaux bayésiens naïf augmenté par un arbre.

Algorithme :

1- Déterminer les fréquences d'apparition de chaque classe à partir de l'ensemble d'apprentissage. Ces fréquences sont calculées comme suit:

$$p(C_i) = \frac{N(C_i)}{N} \qquad et \qquad \sum_{i=1} p(C_i) = 1$$

Sachant que : N (C_i) représente le nombre d'image ou objet appartenant a la classe C_i.

Le N est le nombre total d'images d'apprentissage.

2- calculer la moyenne et la variance de chaque attribut par rapport a ses parents, dans le cas de TAN chaque attribut a un seul parent continu (un autre attribut) et un parent discret (classe).

$$p\left(X_i \mid pa(X_i), C\right) = N\left(m_{i|c}, v_{i|c}\right),$$

$$m_{i|c} = \mu_{i|c} + \sum_{j=1}^{ni} \beta_{ij|c}\left(X_j - \mu_{j|c}\right),$$

$$v_{i|c} = \sigma_{i/c}^2, \qquad \text{et} \qquad \beta_{ji|c} = \frac{\sigma_{ij|c}}{\sigma_{j|c}^2}$$

Avec X_j est le parent de Xi.

Fig.3.10 Algorithme d'apprentissage de paramètres du TAN

Les termes nuls de la diagonale expliquent l'absence de l'information mutuelle $IM(X_i,C)$ entre le même nœud. Le calcul de l'information mutuelle entre les attributs et la classe pour choisir la racine donne les résultats présentés au tableau.3.4.

Tableau.3.4 Information mutuelle entre les attributs et la classe

$IM(X_i,C)$	**S**	**AB**	**ab**	**P**
Classe	0.0994	0.0126	0.0041	0.2038

Notons que le maximum de $IM(X_i, C)$ *est de* 0.2038 correspondant au paramètre périmètre *P* qui, constitue ainsi, le nœud périmètre constitue la racine. Ceci nous permet d'obtenir la structure de la figure (3.11).

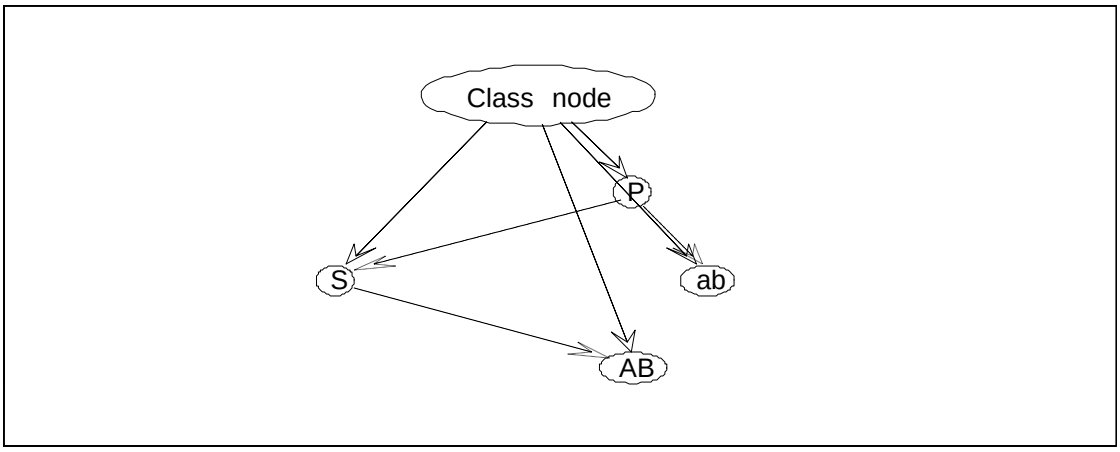


Fig.3.11 TAN obtenu avec les paramètres de forme

Cette structure se présente sous une forme arborescente où chaque attribut a deux parents: la classe et un autre attribut continu alors que le nœud racine *B* ne possède juste qu'un seul parent à savoir la classe. Par ailleurs, le nombre d'arcs est bien égal à trois (représentés en pointillé). Une fois la structure définie, les relations de causalités sont déterminées, ce qui permet de calculer la matrice des poids représentant les coefficients de corrélations entre les nœuds dépendants. La matrice de poids *WT* obtenue de chaque attribut par rapport à sont parent pour chaque classe en utilisant la formule suivante :

$$\beta_{ji|c} = \frac{\sigma_{ij|c}}{\sigma^2_{j|c}} \qquad (3.7)$$

Est donnée par la matrice suivante :

$$WT = \begin{bmatrix} 0.8185 & 0.8246 & 0.8375 \\ 0.5255 & 0.4944 & 0.5477 \\ 0.2287 & 0.0839 & 0.2169 \\ 0 & 0 & 0 \end{bmatrix}$$

Les termes nuls de la quatrième ligue expliquent l'absence du parent pour cet attribut. Effectivement on voit bien que dans la structure résultante, l'attribut *P* est sans parent continu. Une fois cette matrice obtenue, on calcule toutes les moyennes et les variances conditionnelles avec l'équation donnée dans l'algorithme précèdent.

$$m_{i|c} = \mu_{i|c} + \sum_{j=1}^{ni} \beta_{ij|c} \left(X_j - \mu_{j|c} \right) \tag{3.8}$$

Une fois toutes les moyennes et les variances de tous les attributs calculées par rapport à leurs parents pour toutes les classes, le réseau se trouve ainsi quantifié et il peut être utilisé pour classer de nouveaux objets, étant donné leurs nouvelles valeurs d'attributs en utilisant l'équation suivante :

$$P(c_i / x) = P(c_i) \prod_{j=1}^{n} P(x_j / Pa(x_j), c_i) \tag{3.9}$$

Avec :

$$p\left(X_i \mid pa(X_i), C\right) = N\left(m_{i|c}, v_{i|c}\right)$$

Cette équation permet de calculer la probabilité de tous les attributs par rapport à chaque classe autrement l'inférence du TAN. La bonne classe sera celle correspondant à la probabilité est la plus grande. La matrice de confusion en utilisant les paramètres de formes, pour le réseau bayésien naïf augmenté par arbre, est donnée au tableau 3.5.

Tableau.3.5 Matrice de confusion en utilisant les paramètres de forme

	Alp	**Mar**	**Ced**	**Rejet**
Alp	**44**	6	0	0
Mar	2	**48**	0	0
Ced	0	8	**42**	0

En utilisant le réseau bayésien naïf augmenté par un arbre, la matrice de confusion nous montre, qu'ici six grains de pollen de pin d'Alep sont confondus avec le pin maritime. Huit grains du cèdre sont chevauchés avec le pin maritime et en note le même chevauchement pour les grains du pin maritime avec ceux du réseau bayésien naïf.

7.3 Utilisation du réseau bayésien naïf augmenté par forêt

La structure du réseau bayésien naïf augmenté par forêt (FAN) s'obtient par filtrage des liens qui ont des informations mutuelles conditionnelles inférieures à l'Information Mutuelle Conditionnelle Moyenne *IMC* qui est utilisée comme seuil. Ainsi toute information mutuelle conditionnelle inférieure à l'*IMC* moyenne est éliminée selon l'algorithme de la figure (3.12).

Algorithme d'apprentissage de la structure de FAN
Entrée :- L'ensemble d'apprentissage.
- RB Naïf augmenté par un arbre (TAN)
Sortie :- RB Naïf augmenté par un foret (FAN)
Algorithme :

1- Calculer l'information mutuelle conditionnelle pour chaque couple d'attributs donnée par :

$$I_P(X_i; X_j \mid C) = -\frac{1}{2}\sum_{c}^{C} P(c)\log\left(1-\rho_c^2\left(X_i, X_j\right)\right)$$

Avec $\rho_c\left(X_i, X_j\right)=\dfrac{\sigma_{ij|c}}{\sqrt{\sigma_{i|c}^2\sigma_{j|c}^2}}$ et $i \neq j$

2- Calculer l'information mutuelle conditionnelle moyenne définie par l'équation suivante :

$$\mathrm{Im}oy = \frac{\sum_i \sum_j I(Xi, Xj / C)}{n(n-1)} \quad \text{avec} \quad i \neq j$$

3- Construire un graphe complet et non orienté entre les attributs ;
4- Pondérer $I_P(X_i; X_j \mid C)$ sur l'arc (X_i, X_j) ;
5- Calculer l'information mutuelle entre chaque attribut et la classe, IM (X_i, C) pour le choix de la racine
6- Transformer l'arbre non orienté résultante en un arbre orienté en mettant X_{racine} comme racine.
7- Supprimer les liens dirigés qui ont le poids des informations mutuelles Conditionnelles au-dessous de l'information mutuelle conditionnelle moyenne (I_{moy}).
7- Ajouter le nœud C et un arc de C vers chaque nœud X_j.

Fig.3.12 Algorithme d'apprentissage de la structure du FAN

Comme pour le réseau naïf, l'apprentissage des paramètres dans le cas du FAN est basé sur un calcul de fréquences et les paramètres de la loi normale

de gauss comme illustré dans l'algorithme d'apprentissage de paramètres de TAN En appliquant l'algorithme précédent, le calcul de l'*IMC* moyenne donne la valeur I_{moy} = 0.9326 en utilisant l'expression suivante :

$$\mathrm{Im}oy = \frac{\sum_i \sum_j I(Xi, Xj/C)}{n(n-1)} \quad \text{avec} \quad i \neq j \tag{3.10}$$

Apres le calcul de cette moyenne, les valeurs qui sont au-dessous de cette valeur sont filtrées. La matrice du score obtenue est la suivante:

$$IMC = \begin{bmatrix} 0 & 1.1221 & 0 & 2.7264 \\ 1.1221 & 0 & 0 & 1.1080 \\ 0 & 0 & 0 & 0 \\ 2.7264 & 1.1080 & 0 & 0 \end{bmatrix}$$

Ceci nous permet d'obtenir le FAN sous la forme de la structure de la figure (3.13).

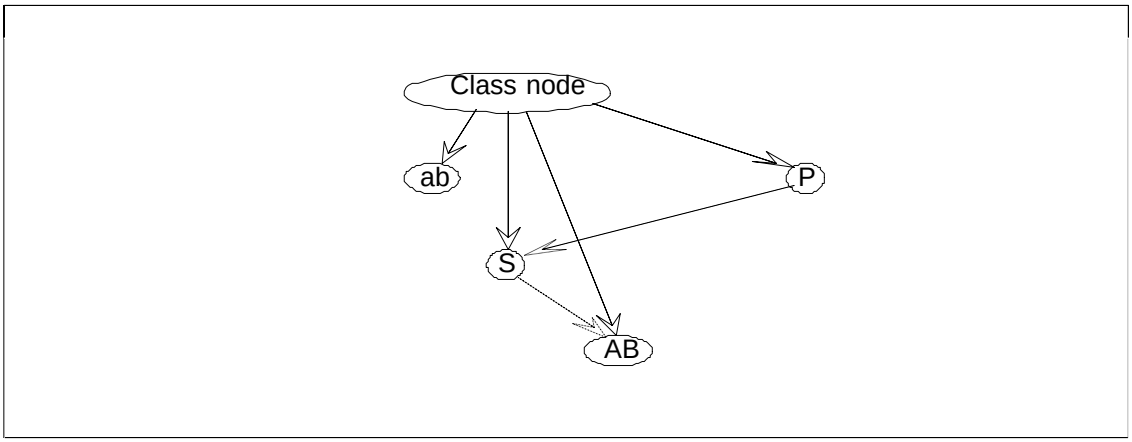


Fig.3.13 FAN obtenu avec les paramètres de formes

Le résultat confirme que, seules, quelques valeurs de probabilité sont significatives tandis que quelques liens peuvent ne pas être nécessaires d'exister dans le TAN. Une fois la structure définie, là aussi, on procède au calcul des paramètres. Pour ce faire, on détermine la matrice de poids de chaque attribut par rapport à son parent en utilisant l'équation (3.7) puisque même ici chaque nœud peut avoir au maximum un parent continu plus le nœud classe.

$$WF = \begin{bmatrix} 0.8185 & 0.8246 & 0.8375 \\ 0.5255 & 0.4944 & 0.5477 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Dans cette matrice, le nombre de colonnes représente le nombre de classes alors que les ligues, comme pour le TAN, correspondent aux coefficients de corrélations entre les attributs et leurs parents. Les valeurs nulles de la troisième et quatrième ligue correspondent à l'absence de parent pour les attributs *ab* et *P*. Une fois cette matrice obtenue, nous calculons toutes les moyennes et les variances conditionnelles avec l'équation (3.8). Après le calcul de toutes les moyennes et les variances de tous les attributs par rapport à leurs parents pour toutes les classes, les nouveaux objets, étant donné leurs nouvelles valeurs d'attributs, sont classées en utilisant l'équation (3.9). La bonne classe est celle correspondant à la probabilité est la plus grande. La matrice de confusion en utilisant les paramètres de formes, pour le réseau bayésien naïf augmenté par foret, est donnée au tableau 3.6.

Tableau.3.6 Matrice de confusion en utilisant les paramètres de forme

	Alp	**Mar**	**Ced**	**Rejet**
Alp	**46**	4	0	0
Mar	2	**48**	0	0
Ced	0	4	**46**	0

Cette matrice de confusion nous montre, dans ce cas, que, seulement, quatre grains du cèdre sont confondus avec le pin maritime contrairement aux deux autres modèles où l'on avait 5 et 8 respectivement avec le RB et le TAN. La confusion est la même pour le modèle TAN en ce qui concerne deux grains du pin maritime confondus avec du pin d'Alep. Par contre, le modèle FAN donne de meilleurs résultats que le modèle RB en ce qui concerne la confusion des grains du pin d'Alep avec ceux du pin maritime puisque six grains du pin d'Alep sont confondus avec ceux du pin maritime alors que avec le modèles FAN, ce chiffre est de 4. Ainsi, nous constatons remarquons une nette supériorité du FAN par rapport aux TAN et RBN.

8. Reconnaissance à partir des paramètres de texture

En suivant la même démarche utilisant les paramètres de forme, et en considérant les bases de paramètres de texture d'apprentissage et de test, nous avons utilisé les trois modèles de réseaux bayésiens comme précédemment.

8.1 Utilisation du réseau bayésien naïf

Comme pour les paramètres de forme, pour les 150 images des trois espèces de pollen représentant l'ensemble d'apprentissage, le RBN correspondant aux paramètres de texture utilisés est donné à la figure (3.14).

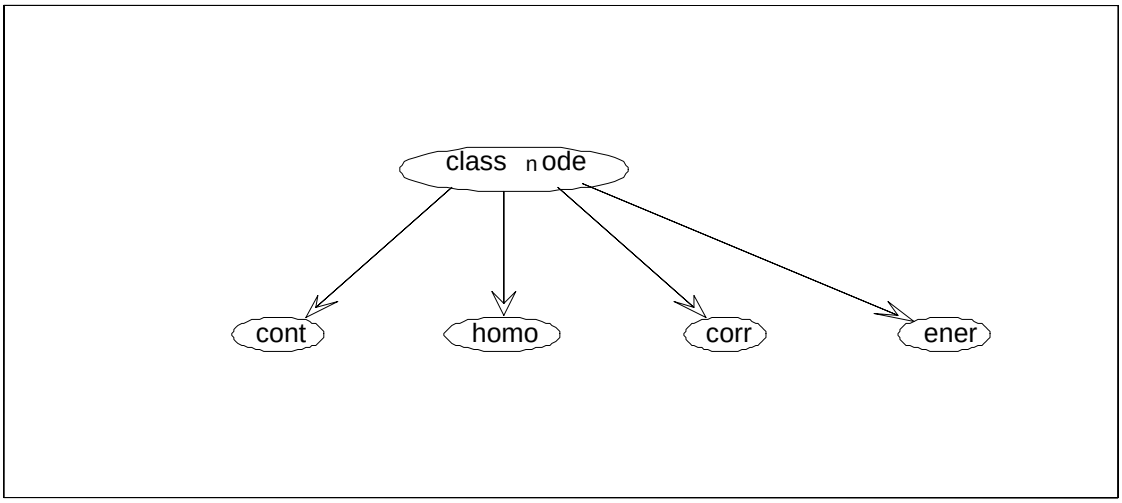


Fig.3.14 RBN obtenu en utilisant les paramètres de texture

La matrice de confusion en utilisant les paramètres de texture, pour le réseau bayésien naïf est donnée au tableau 3.7.

Tableau.3.7 Matrice de confusion en utilisant les paramètres de texture

	Alp	**Mar**	**Ced**	**Rejet**
Alp	**46**	4	0	0
Mar	2	**46**	2	0
Ced	0	3	47	0

Cette matrice de confusion montre que quatre grains de pollen du pin d'Alep sont confondus avec les grains du pin maritime. Deux gains de pin maritime sont confondus avec le pin d'Alep et deux autres sont confondus avec le

cèdre. Trois grains seulement du pollen du cèdre sont confondus avec le pin maritime.

8.2 Utilisation du réseau bayésien naïf augmenté par un arbre

En suivant la même procédure et le même algorithme présentés en 7.2, la structure su TAN obtenue du TAN avec les paramètres de texture utilisés est donnée en figure (3.15).

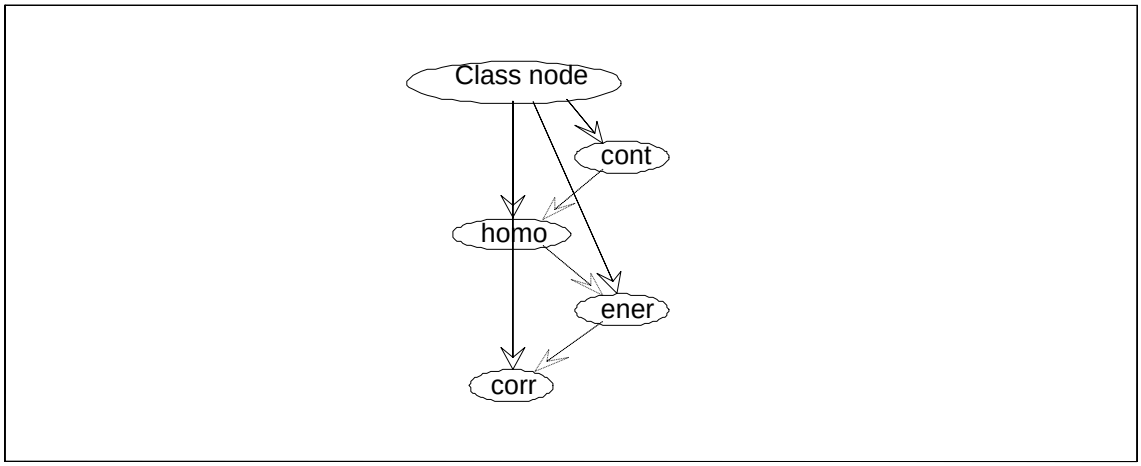


Fig.3.15 Réseau bayésien augmenté par arbre obtenu avec paramètres de texture

La matrice de confusion dans ce cas est donnée au tableau 3.8.

Tableau.3.8 Matrice de confusion en utilisant les paramètres de texture

	Alp	**Mar**	**Ced**	**Rejet**
Alp	**44**	6	0	0
Mar	6	**43**	1	0
Ced	0	0	**50**	0

Cette matrice de confusion montre que six grains de pollen du pin d'Alep sont confondus avec les grains du pin maritime. De son côté, pour le pin maritime, six grains sont affectés au pin d'Alep et un grain est confondu avec ceux du cèdre. Par contre, pollen du cèdre est la totalement reconnu.

8.3 Utilisation du réseau bayésien naïf augmenté par foret

La structure du modèle du FAN est obtenue en suivant les mêmes étapes décrites au paragraphe 7.3. Cette structure en montrée à la figure (3.16).

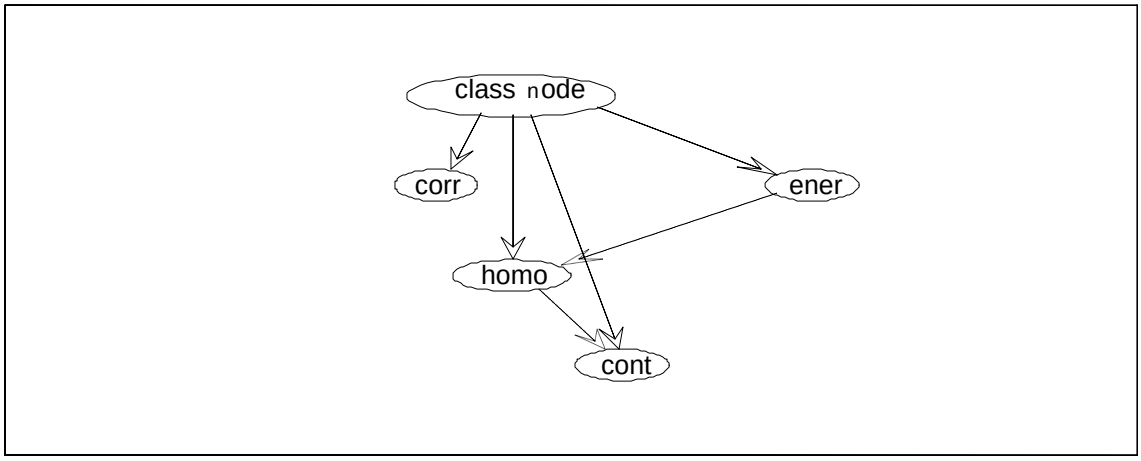


Fig.3.16 Réseau FAN obtenu avec les paramètres de texture

La matrice de confusion obtenue est donnée par le tableau 3.9.

Tableau.3.9 Matrice de confusion en utilisant les paramètres de texture

	Alp	**Mar**	**Ced**	**Rejet**
Alp	**46**	0	4	0
Mar	2	**46**	2	0
Ced	0	1	**49**	0

Pour le réseau FAN, en utilisant la texture, La matrice de confusion montre que quatre grains de pollen du pin d'Alep sont affectés au cèdre, deux gains de pin maritime, au le pin d'Alep et deux grains, au cèdre. Par ailleurs, un seul grain cèdre est affecté au pin maritime.

9. Association des paramètres de forme et de texture

9.1 Utilisation du réseau bayésien naïf

La structure de RBN connue est montrée à la figure (3.17) et le tableau 3.10 donne la matrice de confusion.

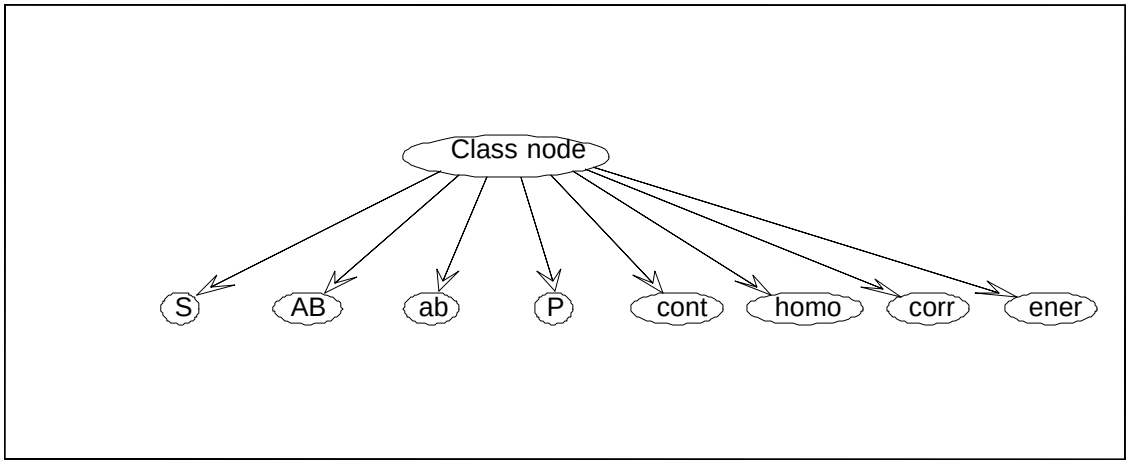


Fig.3.17 RBN obtenu avec association de paramètres

Tableau.3.10 Matrice de confusion en utilisant l'association de paramètres

	Alp	**Mar**	**Ced**	**Rejet**
Alp	**46**	4	0	0
Mar	2	**46**	2	0
Ced	0	1	**49**	0

Cette matrice montre que quatre grains de pollen du pin d'Alep sont confondus avec ceux du pin maritime, deux gains de pin maritime, avec ceux du pin d'Alep et un grain, avec ceux du cèdre. Pour le pollen du cèdre, un grain seulement est affecté à la classe des grains du pin maritime.

9.2 Utilisation du réseau bayésien naïf augmenté par un arbre

De la même manière que dans les cas précédents, la matrice de score (information mutuelle) est obtenue est. Cette matrice étant symétrique, seule la moitié de cette matrice est donnée telle que :

$$IMC = \begin{bmatrix} 0 & & & & & & & \\ 1.1221 & 0 & & & & & & \\ 0.2520 & 0.1187 & 0 & & & & & \\ 2.7264 & 1.1080 & 0.2686 & 0 & & & & \\ 0.1446 & 0.1088 & 0.0973 & 0.1480 & 0 & & & \\ 0.4798 & 0.4063 & 0.2206 & 0.5478 & 0.5519 & 0 & & \\ 0.0635 & 0.0666 & 0.0476 & 0.0713 & 0.1689 & 0.0351 & 0 & \\ 0.7744 & 0.6001 & 0.5544 & 0.8039 & 0.1899 & 0.5225 & 0.1899 & 0 \end{bmatrix}$$

La structure du réseau obtenu est donnée à la figure (3.18).

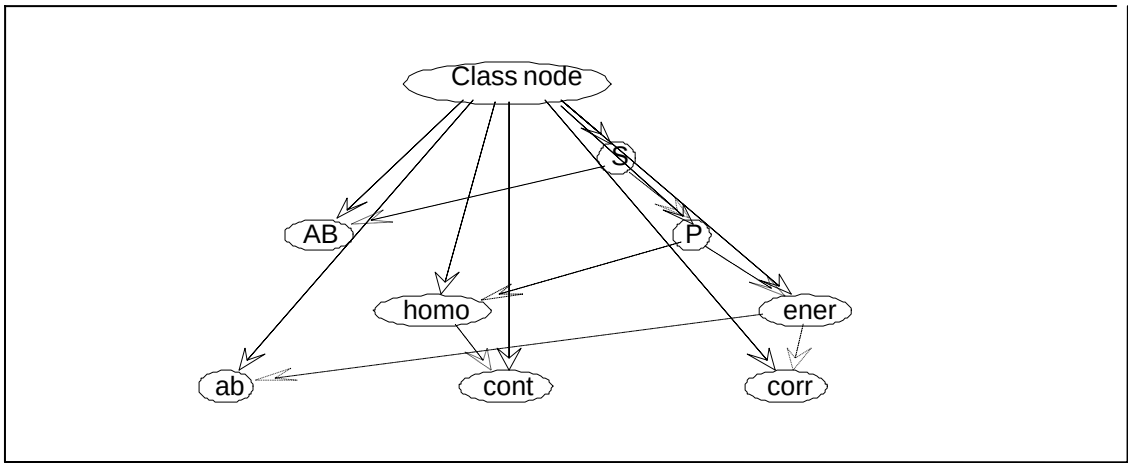


Fig.3.18 TAN obtenu avec l'association des paramètres de forme et de texture

La matrice de confusion en obtenue est donnée au tableau 3.11.

Tableau.3.11 Matrice de confusion en utilisant l'association de paramètres

	Alp	**Mar**	**Ced**	**Rejet**
Alp	**44**	6	0	0
Mar	2	**45**	3	0
Ced	0	2	**48**	0

D'après cette matrice, on constate que six grains de pollen du pin d'Alep sont confondus avec les grains du pin maritime, deux gains de pin maritime sont confondus avec ceux du pin d'Alep et trois autres, avec ceux du cèdre. En ce qui concerne le cèdre, deux de des grains sont confondus avec ceux de pin maritime.

9.3 Utilisation du réseau bayésien naïf augmenté par foret

La figure (3.19) montre la structure obtenue pour le FAN

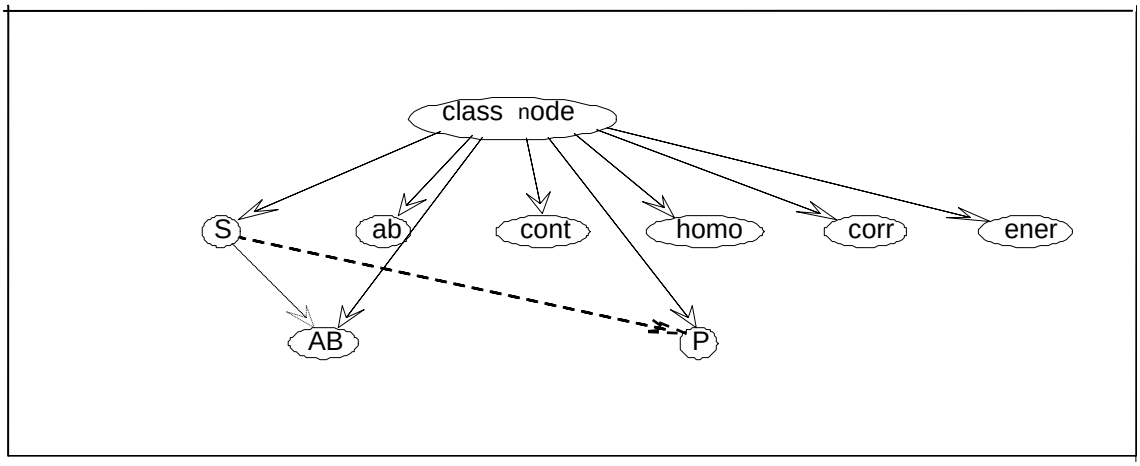


Fig.3.19 FAN obtenu avec association des paramètres de forme et de texture

Le tableau 3.12 donne la matrice de confusion.

Tableau.3.12 Matrice de confusion en utilisant l'association de paramètres.

	Alp	**Mar**	**Ced**	**Rejet**
Alp	**46**	4	0	0
Mar	2	**47**	1	0
Ced	0	0	**50**	0

D'après cette matrice, nous constatons que tous les grains du cèdre sont bien classés alors que quatre de pin d'Alep sont confondus avec ceux de pin maritime, deux de pin maritime avec ceux de pin d'Alep et un avec ceux du cèdre.

10. Récapitulation des résultats et comparaison

Les résultats de reconnaissance en utilisant les paramètres de formes, de texture et du mélange de ces deux types de paramètres sont respectivement donnés aux tableaux 3.13, 3.14 et 3.15.

Tableau 3.13 Résultats de classification en considérant que la forme

	Classe1			Classe2			Classe3			Moyenne		
	V	F	R	V	F	R	V	F	R	V	F	R
RBN	**92**	8	0	**96**	4	0	90	10	0	92.66	7.33	0
TAN	88	12	0	96	4	0	84	16	0	89.34	10.66	0
FAN	92	8	0	96	4	0	92	8	0	93.34	6.66	0

Tableau.3.14 Résultats de classification en considérant que la texture

	Classe1			Classe2			Classe3			Moyenne		
	V	F	R	V	F	R	V	F	R	V	F	R
RBN	92	8	0	92	8	0	94	6	0	92.66	7.33	0
TAN	88	12	0	86	14	0	100	0	0	91.34	8.66	0
FAN	92	8	0	92	8	0	98	2	0	94	6	0

Tableau.3.15 Résultats de classification avec l'association des paramètres

	Classe1			Classe2			Classe3			Moyenne		
	V	F	R	V	F	R	V	F	R	V	F	R
RBN	92	8	0	92	8	0	98	2	0	94	6	0
TAN	88	12	0	90	10	0	96	4	0	91.34	8.66	0
FAN	92	8	0	94	6	0	100	0	0	95.34	4.66	0

Pour chacune des expérimentations effectuées, l'évaluation de l'efficacité de classification est basée sur le taux de bonne classification (V), le taux de mauvaise classification (F) et le taux de rejet (R) de chaque classe, des instances appartenant à l'ensemble test définie par:

- Le nombre des instances appartenant à la classe C_i et qui sont classée dans la classe C_i devisé par le nombre total des instances de la classe C_i, représente le taux de bonne classification.
- Le nombre des instances appartenant à la classe C_i et qui sont classée dans la classe C_j devisé par le nombre total des instances de la classe C_i représente le taux de mauvaise classification.
- Le nombre des instances appartenant a la classe C_i et qui ne sont pas classées ni dans la classe C_i ni dans la classe C_j représente le taux du rejet.

Le tableau 3.13 qui correspond aux résultats de la classification en utilisant que les paramètres de formes, montre que le FAN donne de meilleurs résultats avec un pourcentage de 93.34% de bonne classification. Puis vient le RBN avec un pourcentage de 92.66% et ce même avec la violation de liaison qui peuvent être existées entre les différents nœuds. Tandis que pour le TAN, les 89.33% de bonne classification peut être expliqué par la présence de liens inutiles dans le réseau qui peuvent entraîner de faux résultats.

Les résultats de la classification en utilisant les paramètres texture pour les trois modèles considérés sont résumés dans le tableau 3.14. On peut constater qu'une fois de plus, l'amélioration obtenue par le FAN avec un pourcentage de 94% de bonne classification. La aussi le RBN malgré sa simple structure donne un pourcentage de 92.66% de bonne classification,

ensuite vient le TAN avec 91.34% de bonne classification moins que le FAN et le RBN, étant donné l'existence de liens en plus entre les nœuds.
Le tableau 3.15 résume les résultats de la classification, en associant les deux paramètres, et ce pour les trois modèles testés. La aussi le FAN prend le dessus avec un pourcentage de 95.34% de bonne classification. Puis vient le RBN en deuxième position avec un pourcentage de 94% mieux que le TAN avec un pourcentage de 91.34%.

11. Conclusion

D'après ce chapitre, on peut conclure que l'étude de dépendances entre les attributs a une grande influence sur les résultats de classification d'où l'intérêt de bien choisir les liens c'est-à-dire choisir les liens qui présentent de fortes dépendances entre les attributs qui influent positivement sur les résultats de classification. En effet, un choix de liens bien étudié scientifiquement implique une bonne discrimination entre les classes, ce qui induit certainement une nette amélioration dans le taux de bonne classification, ce qui est le cas de réseau bayésien naïf augmenté par une foret. On peut aussi conclure que pour une bonne discrimination entre les différentes classes de pollen s'obtient par l'association de plusieurs paramètres, dans notre cas en 'a enregistré un taux élevé de bonne classification en associant les paramètres de formes et ceux de la texture.

Conclusion générale

Le théorème de Bayes énoncé au 18ème siècle s'est révélé être un outil très puissant de décision ou d'aide à la décision avec la révolution de l'informatique. Par une approche statistique et probabiliste des données, les réseaux bayésiens permettent d'établir un formalisme des connaissances incertaines, ce qui les place au cœur du domaine l'intelligence artificielle. Plusieurs domaines d'applications sont intéressés par ce type de représentation et de nombreux travaux sont réalisés, preuve de l'intérêt porté par la communauté scientifique à ce type de réseaux.

L'apport des réseaux bayésiens dans les problèmes de classification et la reconnaissance de formes a été notre objectif en utilisant trois variantes des réseaux bayésiens en utilisant les paramètres de forme, les paramètres de texture et les types de paramètres ensemble réunis. En premier lieu les réseaux bayésiens naïfs RN qui malgré leur structure simple ont donnés de très bons résultats. Puis vient les réseaux bayésiens naïfs augmentés par un arbre TAN. En effet, l'hypothèse d'indépendance entre les attributs est généralement fausse. A cet effet, nous avons vu qu'il l existe différentes techniques pour assouplir cette hypothèse. En troisième lieu, les réseaux bayésiens naïfs augmentés par une forêt FAN qui, bien qu'ils soient assez connus dans les problèmes de classification, n'ont pas été explorés suffisamment dans l'imagerie. Les résultats obtenus ont montré une nette amélioration des résultats en utilisant les réseaux FAN par rapport aux réseaux RN et TAN, notamment on associant les paramètres de formes et de texture.

Enfin et après cette approche statique, pour aller plus loin, il pourrait être intéressant de se pencher sur les réseaux bayésiens dynamiques. Ceux ci sont une répétition du réseau classique dans lesquels on rajoute un lien

causal d'un pas de temps à l'autre. Ils contiennent chacun un certain nombre de variables aléatoires représentant les observations et les états cachés du processus. Le temps ici est discret et chaque unité de temps représente une nouvelle observation. La complexité différentielle des réseaux bayésiens dynamiques est évidement bien plus élevé que celle des réseaux statiques.

References

[1]: A.D. Brink: Using spatial information as an aid to maximum entropy image threshold. Recognition letters.1996, Vol. 17, pp. 29-36.

[2]: Y. Qiao et al: Thresholding based on variance and intensity contrast. Pattern Recognition Letters. 2007, Vol. 40, pp. 596-608.

[3]: Kittler, J. and Illingworth, J: Minimum error thresholding. Pattern Recognition 1986, Vol.19, 1, pp. 38-52.

[4]: A. Nakib, H. Oulhadj et P. Siarry : Segmentation optimale d'images par optimisation multicritère et dérivation non entière. Congrès de la société Française de Recherche Opérationnelle et d'Aide à la décision (ROADEF'07). pp. 337-338, Février 2007, Grenoble.

[5]:H. Cheng et Y. Sun: A hierarchical approach to color image segmentation using homogeneity. IEEE Trans. on Image Processing. 2000, Vol. 9, 12, pp. 2071-2082.

[6]: N. Otsu: A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. On Syst., Man and Cyb. 1979, Vol. 9, 1, pp. 62-66.

[7]: F. Zhang, S. Q. Liu, D. B. Wang & W. Guan. Review article: Air- craft recognition in infrared image using wavelet moment invariants. Image Vision Comput., vol. 27, no. 4, pages 313–318, 2009.

[8]: J. Revaud, G. Lavoue & A. Baskurt: Improving Zernike Moments Comparison for Optimal Similarity and Rotation Angle Retrieval. IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 4, pages 627– 636, 2009.

[9]: R.Bouillot : cours de traitement numérique de l'image. Édition Dunod.2005.

[10]: M. J. Swain et D. H. Ballard: Color indexing. Int. J. Comput. Vision, vol. 7, no. 1, pages 11–32, 1991.

[11]: M. Mignotte: Segmentation by Fusion of Histogram-Based K-Means Clusters in Different Color Spaces. IEEE Trans. Image Processing, vol. 17, no. 5, pages 780–787, 2008.

[12]: R. M. Haralick: Statistical and structural approaches to texture. Proceedings of the IEEE, vol. 67, no. 5, pages 786–804, 1979.

[13]: P.J. Li, T. Cheng et J.C. Guo : Multivariate Image Texture by Multivariate Variogram for Multispectral Image Classification. PhEngRS, vol. 75, no. 2, pages 147–158, 2009.

[14]: H.Majdoulayne : Extraction de caractéristiques de texture pour la classification d'images satellites. Thèse de doctorat de l'université de Toulouse. Année 2009

[15]: S.Mavromatis : Analyse de texture et visualisation scientifique. Thèse de doctorat à l'Université de la Méditerranée. Spécialité informatique. Année 2001

[16]: M. Unser : Sum and difference histograms for texture classification. IEEE Transactions on pattern analysis and machine intelligence 8(1): 118-125 (1986).

[17]: O.Bouhaddou : Analyse en composantes principales et interpolation de processus », thèse de spécialité. Grenoble.1984.

[18]: K.WAGSTAFF, C.CARDIE: Constrained K-means clustering with background knowledge. proceedings of the eighteenth international conference on machine learning 2001.p.577-585.

[19]: N.OUARAB, Y.SMARA et J.P.RASSON : Utilisation de méthodes de classification hiérarchique pour une classification supervisée d'images satellitaires. Dix septième colloque GRETSI, vannes. 13-17. September 1999.

[20]: T.M. Cover et P.E. Hart: Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13:21–27, 1967.

[21]: A. Vellido, P. Lisboa et K. Meehan: Segmentation of the online shopping market using neural networks. Expert Systems with Applications, 17, 303–314.1999.

[22]: P. LARAY, O. François : Réseaux bayésiens pour la classification, méthodologie et illustration dans le cadre du diagnostic médical. PhEngRS, vol. 80, no. 3, pages 140–150, 2002.

[23]: N.B.Amor, S.Benferhat et Z.Elouedi : Réseaux bayésiens naïfs et arbre de décision dans les systèmes de détection d'intrusions. RSTI-TSI. Volume 25 – nÆ2/2006, pages 167 à 196.

[24]: A. Cornuéjols : Une nouvelle méthode d'apprentissage : Les SVM. Séparateurs à vaste marge ». Université de Paris Sud, Orsay. Numéro 51 - juin 2002

[25]: S.Verron. Diagnostic et surveillance des processus complexes par réseaux bayésiens. Thèse doctorat. Université d'Angers. 2007.

[26]: F.Pernkopf : Bayesian network classifiers versus sélective K-nn classifier pattern recognition. Thèse de doctorat de l'Université de François Rabelais de Tours. Année 2005.

[27]: M.Perzyk, R.Biernacki et A. Kochanski: modeling of manufacturing processes by Learning systems: The naive Bayesian classifier versus artificial neurol networks. Journal of materials processing technology 2005.

[28]: Ann.Becker, Patrick Naim : les réseaux bayésiens : modèles graphiques de connaissance.Eyrolles.1999.

[29]: K.Hallouli: Reconnaissance de caractères par method markoviennes et réseaux bayesiens. Thèse de doctorat de Telecom Paris. École nationale supérieure des telecommunications. Spécialité Signal et Images. Année 2004

[30] : P. Leray et O. François : Réseaux Bayésiens pour la classification méthodologie et illustration dans le cadre du diagnostic médical. INSA Rouen / PSI, FRE CNRS 2645. BP 08 - Av. de l'Université.76801, St Etienne du Rouvray Cedex. 2002

[31]: D.Bellot : Fusion de données avec des réseaux bayésiens pour la modélisation des systèmes dynamiques et son application en télémédecine. Thèse doctorat. Université de Bruxelles. Année 2002.

[32]: V.Labatut : Réseaux causaux probabilistes à grande échelle : un nouveau formalisme pour la modélisation du traitement de l'information cérébrale. Thèse doctorat de l'Université de Aix-Marseille. Année 2003.

[33]: Laurence Grammont : cours de probabilité 2eme année d'économie et de gestion semestre 2. 2004.

[34]: L.Smail : Algorithmes pour les réseaux bayésiens et leurs extensions. Thèse doctorat de l'Université de Polytech Nantes. Année 2004.

[35]: Eduardo Sanchez Soto : Réseaux bayésiens dynamiques pour vérification du locuteur. Thèse doctorat 2005.

[36]: Pascal Cheung- Mon- Chan : Réseaux bayésiens et filtres particuliers pour l'égalisation adaptative et le décodage conjoints thèse doctorat de l'école normale supérieure de Cachan. Spécialité mathématique. Année 2006.

[37]: Tom .Mitchell: Generative and discriminative classifier: Naive bayes and logistic regression. Machine learning. Draft 2010.

[38]: Y.Xin, Z.Zhaobao, Z.Haitao, Y. Zhiwei: Texture classification of aerial image based on bayesian network augmented naive bayes. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B7. Beijing 2008

[39]: M.A.mahjoub, K.Jayech : indexation de structure de documents par réseaux bayésiens.

[40]: F.Dress: Probabilités et statistiques de A à Z. 500 définitions, formules et tests d'hypothèse. Edition Dunod.paris.2004.

[41]: Sabine Barrat : Modèles graphiques probabilistes pour la reconnaissance de formes. Thèse doctorat de l'Université Nancy 2. Spécialité informatique. Année 2009.

[42]: M.A.mahjoub, K.Jayech : indexation de structure de documents par réseaux bayésiens.

[43]: Aabid Shariff, Ahmed Bakan: Tree augmented naive bayesian classifier with feature selection for FMRI data.

[44]: C. Aaron : Algorithme EM et classification non supervisée. Thèse de doctorat de Université Paris I. année 2001

[45]:O.Francois, P.Leray : étude comparative d'algorithmes d'apprentissage de structure dans les réseaux bayésiens. Proceedings of the IEEE, vol. 60, no. 4, pages 586–704, 2004.

[46]: Z. Yun et K. Keong: Improved MDL scores for learning of Bayesian networks. Dans Proceedings of the International Conference on Artificial Intelligence in Science and Technology. AISAT 2004, pp. 98–103.

[47]:W.Tekaya : Problème de l'arbre couvrant de poids minimum probabiliste. Master modélisation, optimisation, décision et organisation. Rapport du mémoire majeure 2007/2008.

[48]: N.Sendier : Introduction à la théorie de l'information. Polytechnique, août 2007.

[49]: O.Mariano, G.Gervais : Kruskal et Prim : manipulation des graphes, 2007.

[50]: A.Cornuéjols. L.Miclet : Apprentissage artificiel. Concepts et algorithmes. Edition : Eyrolles.2002.

[51]: C.Nicolas: Apprentissage de structure des réseaux bayésiens à partir de données incomplètes. Mémoire d'ingénieur en informatique spécialité réseaux et systèmes. 2008

[52]: Camille Séka Kotchi, Véronique Delacroix et Sylvain Piechowiak : Etude de la performance des algorithmes d'inférence dans les réseaux Bayésiens. Laboratoire d'Automatique, de Mécanique et d'Informatique industrielles et Humaines, Université de Valenciennes et du Hainaut Cambrésis.

[53]: V. Mazet : Introduction aux méthodes de Monte Carlo par chaînes de markov. CRAN.CNRS UMR 7039, Université Henri Poincaré, 54506 Vandoeuvre lès nancy Cedex. 22 mai 2003

[54]: K.Tanaka, J. I. Imoune et D.M.Titterington: loopy belief propagation and probabilistic image processing.1999.

[55]: Li, P. & FLENLEY, J.R: Pollen texture identification using neural networks.Grana 38: 59-64, ISSN 0017-3134, 1999.

[56]: P. Bonton, A. Boucher et al: colour image in 2d and 3d microscopy for the automation of pollen rate measurement. In proceeding of 8th ECS and Image Analysis , Bordeaux, France, 2001.

[57]: S.Bouguenna : Diagnostic écologique, mise en valeur et conservation des pineraies de pinus halpensis de la région de Djerma. (Nord-est du parc national de Belezma, Batna. Thèse de magister, spécialité agronomie.2010-2011.

[58] : A.Erchan : Analyse d'images couleur par morphologie mathématique : application à la description, l'annotation et la recherche d'images. Thèse de doctorat, université Louis Pasteur.2008.

[59]:L.Richard : Reconnaissance d'algues toxiques par vision artificielle et réseau de neurones. Mémoire de maîtrise. Université du Québec.2004.

[60] : O. Boukendour : Contribution à la l'automatisation de la détection et de la reconnaissance de pollens. mémoire de magister en automatique, option traitement d'images et reconnaissance de formes à UMMTO. 2011.

[61] : O. Boukendour : Contribution à la l'automatisation de la détection et de la reconnaissance de pollens. mémoire de magister en automatique, option traitement d'images et reconnaissance de formes à UMMTO. 2011.

[63]: R.Haralick. K.Shanmugan et S.Dinstein: Textural features for image classification. IEEE Transactions on systems, Man and cybernetics 3(6):610-621.

[64]: Murphy: the bayes net toolbox for Matlab computing science and statistics.2001

[65]: P.Leray et O.Francois: structure learning package, documentation and experiments. November 2006, version 1.3

[66]: François Olivier: de l'identification de structure des réseaux Bayésiens à la reconnaissance de forme à partir d'information complètes ou incomplètes. Thèse doctorat.2006.

[67]: Patrick Naim, Pierre- Henri Wuillemin, Philippe Leray, Olivier Pourret et Anna Becker : Réseaux bayesiens 2eme édition Eyrolles 2004.

Annexe1: Exemple de EM.

Soit la base d'exemple suivante :

M	F	R
m0	Bad	O
m0	Bad	O
?	Bad	O
m0	Bad	O
?	Bad	N
m0	Ok	O
m1	Bad	O
m1	Bad	N
?	Ok	O
m1	Ok	N
m1	Ok	O
m1	Ok	N
m1	?	O
m1	Ok	N
m2	Ok	N

Pour estimer P (M).

1) initialisation

$p^{(0)}(m)= [1/3 \;\; 1/3 \;\; 1/3]$.

1ère itération [E] :

M	F	R	P(M=m0)	P(M=m1)	P(M=m2)
m0	Bad	O	1	0	0
M0	Bad	O	1	0	0
?	Bad	O	1/3	1/3	1/3
M0	Bad	O	1	0	0
?	Bad	N	1/3	1/3	1/3
M0	Ok	O	1	0	0
M1	Bad	O	0	1	0
M1	Bad	N	0	1	0
?	Ok	O	1/3	1/3	1 /3
M1	Ok	N	0	1	0
M1	Ok	O	0	1	0
M1	Ok	N	0	1	0
M1	?	O	0	1	0
M1	Ok	N	0	1	0
M2	Ok	N	0	0	1
	total		5	8	2

1er itérarion [M] :

$P^{(1)}(m0)=5/15=0.333$

$P^{(1)}(m1)=8/15=0.533$

$P^{(1)}(m2)=2/15=0.133$

2eme itérarion [E] :

M	F	R	P(M=m0)	P(M=m1)	P(M=m2)
m0	Bad	O	1	0	0
M0	Bad	O	1	0	0
?	Bad	O	0.333	0.533	0.133
M0	Bad	O	1	0	0
?	Bad	N	0.333	0.533	0.133
M0	Ok	O	1	0	0
M1	Bad	O	0	1	0

?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O ?● O	?	O	0	1	0